REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	1995	

4. TITLE AND SUBTITLE
Multi-response Nonlinear Mixed Effect for Longitudinal Data Analysis

5. FUNDING NUMBERS

6. AUTHOR(S)
James H. Rutledge III

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
AFIT Students Attending:
University of Colorado, Health Science Center

8. PERFORMING ORGANIZATION REPORT NUMBER
AFIT/CI/CIA
95-024

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
DEPARTNEMT OF THE AIR FORCE
AFIT/CI
2950 P STREET, BDLG 125
WRIGHT-PATTERSON AFB OH 45433-7765

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for Public Release IAW AFR 190-1
Distribution Unlimited
BRIAN D. GAUTHIER, MSgt, USAF
Chief Administration

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

14. SUBJECT TERMS	15. NUMBER OF PAGES 111
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT

MULTI-RESPONSE NONLINEAR MIXED EFFECTS MODELS FOR LONGITUDINAL DATA ANALYSIS

by

JAMES H. RUTLEDGE III

B.A., University of Cincinnati, 1983

M.S., University of Cincinnati, 1989

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Preventive Medicine and Biometrics

1995

ABSTRACT/THESIS INFORMATION

AUTHOR: James H. Rutledge III

TITLE: Multi-response Nonlinear Mixed Effects Models for Longitudinal Data Analysis

RANK: Captain (O-3)

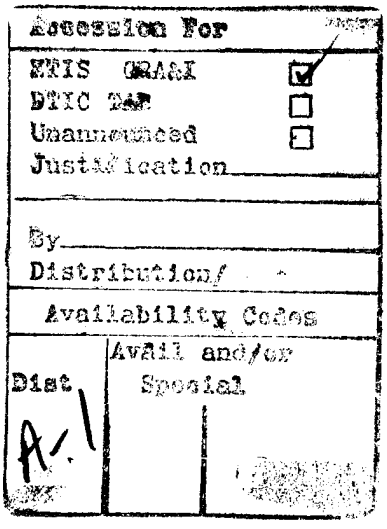
BRANCH: United States Air Force

DATE: 1995

PAGES: 111

DEGREE: Ph.D.

SCHOOL: University of Colorado Health Sciences Center



This thesis for the Doctor of Philosophy degree by

James H. Rutledge III

has been approved for the

Department of

Preventive Medicine and Biometrics

by

Gary O. Zerbe

Richard H. Jones

Date 5/3/95

Rutledge, James H. III (Ph.D., Analytic Health Sciences)

Multi-response Nonlinear Mixed Effects Models for Longitudinal Data Analysis

Thesis directed by Professor Gary O. Zerbe

This thesis presents a method for analyzing multi-response nonlinear longitudinal data. Currently, the methods used to handle multi-response nonlinear longitudinal data usually involve a two stage approach. First, nonlinear functions for each variable are fit to each subject. Second, the parameters from stage one are then used to estimate population average parameters, estimate parameter correlations, conduct hypothesis tests, etc. This two stage approach is often cumbersome because it involves modeling each individual separately. Sometimes the two stage approach is impossible because there might be inadequate data to fit a nonlinear function to certain subjects. This thesis presents a unified approach for fitting multi-response nonlinear mixed effects models (MNLMEM) to longitudinal data. Essentially the nonlinear aspect of the model is handled by Taylor series expansion. Once the model has been "linearized", a multi-response analog of the Laird and Ware model (*Biometrics* **38:** 963-974,1982.) that has been developed by Zucker, Zerbe, and Wu (*Biometrics*, in press) is then applied. In addition, if the errors in the model are additive and the model has been "linearized", it is also possible to use an algorithm discussed by Hocking (*The Analysis of Linear Models,* 1985). Using either approach it is possible to obtain estimates of the fixed effects, variance components, and Fisher's information matrix for both the fixed effects and variance components. This makes it possible to conduct asymptotic hypothesis tests and build asymptotic confidence intervals about functions of the fixed effects and variance

components. The methods are very general and allow for missing and unequally spaced data.

Signed Gary O. Zerbe

Faculty member in charge of thesis

DEDICATION

To my family, especially Joan, Jimmy, and Robert.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to Gary O. Zerbe for the leadership and guidance that he unselfishly provided throughout my Ph.D. program. His patience and mentoring skills have had a direct positive influence on me and I am forever indebted to him for this.

I would also like to thank David A.Young for our many insightful discussions about nonlinear mixed effects models.

The other members of my committee: Richard H. Jones, Philip G. Archer and Woodruff Emlen, to them I express my gratitude.

The other faculty members and administrative support staff have my heart-felt appreciation.

I'd like to thank Colonel Litwhiler and the United States Air Force for providing me with the time, support and, foremost, this incredible opportunity.

Thank you all!

CONTENTS

TABLES

TABLE

FIGURES

FIGURE

CHAPTER I

INTRODUCTION

A good starting place for this thesis will be to explain the title: Multi-response Nonlinear Mixed Effects Models for Longitudinal Data Analysis.

1.1 Longitudinal Data

Longitudinal data are characterized by two features. First, repeated measurements on the same subject over time or dose are collected. Second, each subject can be described with a similar intrasubject model that relates the response variable to time or dose. Figure 1.1 shows a plot of some longitudinal data. The lines are used to connect the measurements for each subject. Longitudinal data analysis methods have widely been used in dose response curves and growth curve analysis. Since there are repeated measurements on the same subject, the observations in longitudinal data are often correlated. Therefore, the usual regression models which assume independence between observations are not appropriate.

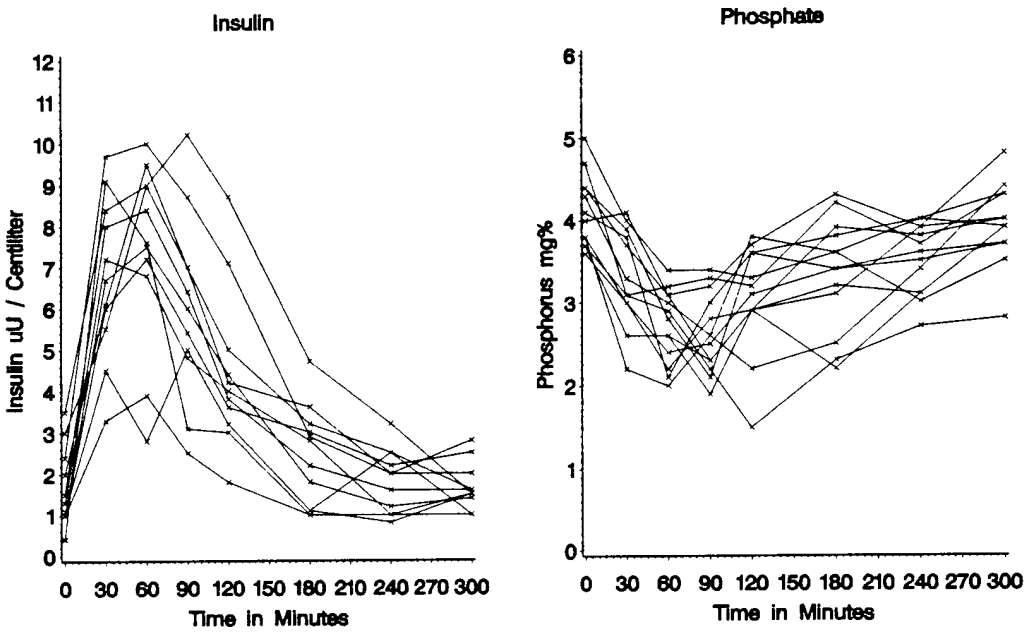
1.2 Multivariate vs Univariate

The "multi-response" part of the title implies that interest lies in more than one response variable. Consider the data shown in Figure 1.1. These data arise from a glucose tolerance test. Blood samples are drawn from each patient at 30 to 60 minute intervals. An insulin level and a phosphate level is then measured for each patient at each of the different time periods. Thus, these data are multivariate since the investigator is interested in both glucose and phosphate, and these values are measured

simultaneously on each patient at each time period. In medical research the multivariate aspects of data are often overlooked or ignored. Typically the data shown in Figure 1.1 would be analyzed using two separate univariate analyses.

Figure 1.1

Insulin and phosphate levels for 13 patients after administration of a glucose tolerance test.



1.3 Nonlinear vs Linear

Any model of the following form is considered to be a linear model:

$$Y = \beta_0 + \beta_1 Z_1 + \cdots + \beta_p Z_p + \varepsilon. \tag{1.1}$$

Here the Z_i can be any function of the explanatory variables $X_1, X_2, ..., X_k$. What distinguishes model (1.1) is that it is linear in the parameters. Some simple examples follow.

Example 1: Consider the simple linear regression case of regressing weight on height. The model would be:

$$Weight = \beta_0 + \beta_1 Height + \varepsilon. \tag{1.2}$$

Example 2: If the relationship between height and weight was not exactly a straight line we could add a quadratic term to the model to account for some of the curvature. The model would be:

$$Weight = \beta_0 + \beta_1 Height + \beta_2 Height^2 + \varepsilon. \tag{1.3}$$

So, even though the previous equation is not a straight line, it is still a linear model because it is linear in the parameters. Any model that is not of the form (1.1) is said to be a nonlinear model. That is, a model is nonlinear if it is not linear in the parameters. Some examples may help to clarify this point.

Example 3: Exponential decay curves are often fit to pharmacokinetic data. For example consider measuring the elimination of hydrogen with the following model:

$$Hydrogen = \alpha \exp\{\beta \cdot time\} + \varepsilon. \tag{1.4}$$

This model is nonlinear in the parameter β. However, by taking the natural logarithm of both sides, the model becomes:

$$\ln(Hydrogen) = \ln(\alpha) + \beta \cdot time + e. \tag{1.5}$$

Now the model appears to be of the form (1.1). We say that this model is intrinsically linear since it can be put into a linear form. In this case, the form of the error is affected.

Example 4: In the area of growth curve analysis, nonlinear models are often employed. These models are typically mechanistic in that they are derived from some assumptions about growth that usually depend on difference equations or differential equations. For example, suppose that we assume that the rate of growth at a particular time, t, is directly proportional to the amount of growth yet to be attained. If α is the maximum growth size and ω is the growth at time t then we have the following differential equation:

$$d\omega/dt = k(\alpha - \omega). \tag{1.6}$$

Here k is the rate constant of the growth curve. If we integrate the above equation we get:

$$\omega = \alpha(1 - \beta \exp\{-k\,t\}). \tag{1.7}$$

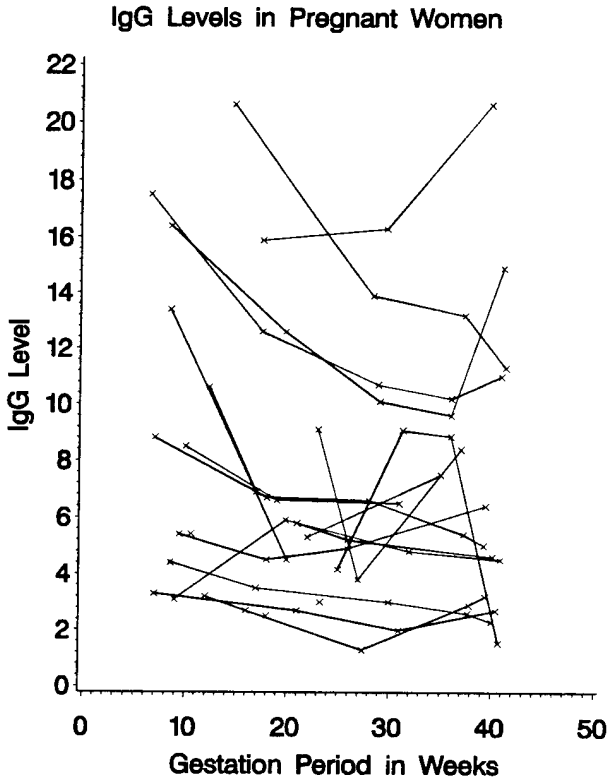
In models such as those described above the parameters are often interpretable. Here the maximum growth is α, and at $t = 0$ the growth curve starts at the point $\alpha(1 - \beta)$. The growth rate k is also readily interpretable (Draper and Smith, 1981).

1.4 Fixed vs Mixed Effects Models

The data shown in figure 1.2 are the immunoglobulin G (IgG) anticardiolipin antibody levels of a group of pregnant women over the course of their pregnancies (Lynch et al., 1994).

Figure 1.2

IgG levels for a sample of 20 pregnant women over the course of their pregnancies.



If a researcher was interested in modeling the trend of IgG over time, one approach would be to use simple linear regression. The model could be:

$$y_{ij} = \alpha + \beta x_{ij} + e_{ij}. \tag{1.8}$$

The y_{ij} is the IgG level for the *i*th women at the *j*th measurement. The x_{ij} is the gestational age of the *i*th women's fetus at the *j*th measurement. The usual regression

assumption is that e_{ij} are *iid* $N(0, \sigma^2)$ [iid, independently identically distributed]. In this situation we call the model a fixed effects model. Since the data in Figure 1.2 contains repeated measures over time on individuals, it is unreasonable to assume that all of the observations are independent. One way around this assumption is to model a line to each individual. That is, for each individual we will have an intercept and slope. Thus the model becomes:

$$y_{ij} = (\alpha + a_i) + (\beta + b_i)x_{ij} + e_{ij}. \tag{1.9}$$

In this model $\alpha + a_i$ is the *i*th woman's intercept and $\beta + b_i$ is the *i*th woman's slope. In addition to assuming that the e_{ij} are iid $N(0, \sigma^2)$ we assume the following:

$$\begin{pmatrix} a_i \\ b_i \end{pmatrix} \sim N(\mathbf{0}, \underset{2\times 2}{\mathbf{D}}). \tag{1.10}$$

Thus in this model we see that the variance of y_{ij} is no longer simply σ^2. This type of model is often called a "stochastic parameter" model, because each parameter in the model is allowed to vary across individuals. In the case of model (1.9) each individual is modeled to allow for individual slopes and intercepts. In this model the α and β are considered fixed effects and the a_i and b_i are considered random effects. Since the model involves both fixed and random effects it is called a mixed model. In general a mixed model is any model where the variance structure is <u>not</u> simply $\sigma^2\mathbf{I}$ (Hocking, 1985).

1.5 Overview

Taken by themselves, each of the previous topics has a rich history in the foundation of statistical knowledge. However, the knowledge base starts to diminish as these topics are overlapped. It is the intersection of multi-response, nonlinear, and mixed effects that make this thesis unique in its undertaking. It is the intent of this thesis to discuss multi-response nonlinear mixed effects models (MNLMEMs). A brief outline for the organization of this thesis follows. Chapter II will be a review of the literature. In Chapter III two MNLMEMs will be introduced. In Chapter IV the methodology will be demonstrated on several data sets. In Chapter V a brief discussion will be offered.

CHAPTER II

LITERATURE REVIEW

It is the intent of this chapter to put the multi-response nonlinear mixed effects model (MNLMEM) into historical perspective with other significant works in the field of longitudinal data analysis.

2.1 Brief Background on Longitudinal Data

As previously stated, the distinguishing factors of longitudinal data are: the variance structure is not simply $\sigma^2\mathbf{I}$ and individuals tend to follow the same general pattern. The earliest article found in the literature that seems to specifically address the problems associated with longitudinal data was written by Wishart (1938). In modeling the growth of bacon pigs over time Wishart (1938) fit quadratic polynomials to each pig. Then he analyzed the resulting regression coefficients across different groups of pigs using the usual ANOVA techniques. The next article found in the literature that seems to address the problems of longitudinal data is authored by Box (1950). Box (1950) describes a method for growth curve analysis that is based on differencing the original data. The differences thus become interpreted as the average growth during successive periods. Often the result of differencing yields the simple compound symmetric covariance structure found in agricultural split-plot designs. In these studies each subject is treated as a "plot" and the multiple measurements per subject are treated as "subplots." Box (1950) goes into several tests for determining if the simple compound symmetric variance structure is appropriate. Box (1950) was among the first to analyze longitudinal

data with multivariate methods that allowed for an unstructured covariance matrix. The mixed effects models would soon replace the multivariate models since they could handle data that were unequally spaced and/or missing. Both of these approaches will be discussed in the following two sections.

2.2 Longitudinal Data via Multivariate Analysis

C.R. Rao (1958, 1959, 1965) played an instrumental role in the development of longitudinal data analysis. His models centered around balanced data, that is, he assumes that each of n subjects are measured at the same q times and no data are missing. Rao uses the traditional two stage approach. First, a growth curve is fit to each individual then an average growth curve is estimated. Rao's model for the ith individual (i=1,...,n) is:

$$\underset{q\times 1}{\mathbf{Y}_i} = \underset{q\times p}{\mathbf{X}}\,\underset{p\times 1}{\boldsymbol{\beta}} + \underset{q\times 1}{\mathbf{E}_i}\ . \tag{2.1}$$

In model (2.1) $\mathbf{Y}_i$ is a data vector for the ith individual. $\mathbf{X}$ is a known design matrix that does not change from individual to individual. $\boldsymbol{\beta}$ is an unknown vector of parameters to be estimated. $\mathbf{E}_i$ is a vector of random errors, where it is usually assumed that:

$$\mathbf{E}_i \sim N(\mathbf{0}, \underset{q\times q}{\boldsymbol{\Sigma}}). \tag{2.2}$$

Rao (1959) showed that the unbiased minimum variance estimator for model (2.1) is:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{S}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{S}^{-1}\bar{\mathbf{Y}} \tag{2.3}$$

where $\mathbf{S}$ is the sample covariance matrix and $\bar{\mathbf{Y}}$ is the mean vector of the data.

Pothoff and Roy (1964) extend model (2.1) by adding a post-matrix to the model. Their generalized model for all n individuals is:

$$\underset{q\times n}{\mathbf{Y}} = \underset{q\times p}{\mathbf{B}}\ \underset{p\times m}{\boldsymbol{\xi}}\ \underset{m\times n}{\mathbf{A}} + \underset{q\times n}{\mathbf{E}}. \tag{2.4}$$

The columns of $\mathbf{Y}$ are mutually independent and the q elements in the columns follow a multivariate normal distribution with expectation $\mathbf{B}\boldsymbol{\xi}\mathbf{A}$ and unknown variance $\boldsymbol{\Sigma}$ ($q \times q$, and positive definite). The matrices $\mathbf{A}$ and $\mathbf{B}$ are respectfully the across and within subjects known design matrices. $\boldsymbol{\xi}$ is a matrix of unknown parameters that needs to be estimated. The generalized model is felt to be more convenient for handling longitudinal data. Pothoff and Roy (1964) propose that estimation in model (2.4) can be accomplished by transforming the original data of model (2.4) as follows:

$$\underset{p\times n}{\mathbf{Y}_1} = (\mathbf{B}'\underset{q\times q}{\mathbf{G}}^{-1}\mathbf{B})^{-1}\mathbf{B}'\mathbf{G}^{-1}\mathbf{Y}. \tag{2.5}$$

Here $\mathbf{G}$ is any matrix such that $(\mathbf{B}'\mathbf{G}^{-1}\mathbf{B})^{-1}$ exists. Under the transformation (2.5) the data can now be analyzed using well known multivariate analysis of variance (MANOVA) techniques (see, e.g. Roy, 1957, Chapter 12). Pothoff and Roy (1964) discuss choices of the $\mathbf{G}$ matrix. They conclude that the 'farther' away $\mathbf{G}$ is from $\boldsymbol{\Sigma}$ the

worse the power of tests will be and the wider the confidence intervals will be. Khatri (1966) found the maximum likelihood estimates of model (2.4). He shows that the maximum likelihood estimate will occur when **G**=**S** (the sample covariance matrix). He also develops the theory necessary to test the following hypothesis:

$$H_o : \underset{c\times p}{\mathbf{F}} \; \underset{p\times m}{\boldsymbol{\xi}} \; \underset{m\times g}{\mathbf{C}'} = \underset{c\times g}{\mathbf{0}}. \tag{2.6}$$

Rao (1965) comments that in Pothoff and Roy's model (2.4) using an arbitrary **G** matrix in the transformation (2.5) does not make use of all the information in the sample unless **G** happened to equal $\boldsymbol{\Sigma}$. Rao (1965) also suggests that it may be beneficial to weight by something other than **S**. He calls this method "adjusting for concomitant variation." Grizzle and Allen (1969) synthesize the work of Pothoff and Roy (1964), Rao (1965), and Khatri (1966) by showing how Rao's "concomitant" variables can be included in model (2.4).

A drawback of the previous models is that they all require that there are no missing data and that all data must be collected at the same points in time or space. In medical research the above requirements are often impossible to achieve. Another disadvantage of these models is that there are a large number of variance components to be estimated if q is large. Therefore, a univariate approach to the problem may be more beneficial. The next section will focus on the analysis of longitudinal data via the univariate mixed effects model.

2.3 Longitudinal Data via the Univariate Linear Mixed Effects Model

Based on the work of Harville (1976, 1977), Laird and Ware (1982) popularized a mixed model that is perfectly geared for longitudinal data analysis. By writing the model in a subject specific format and taking advantage of the fact that subjects are independent of each other they were able to keep the size of the necessary computational matrices to a minimum. This made it possible to implement their model on desktop computers. Their model plays such a significant role in the analysis of longitudinal data that it is worth examining. It will be assumed that there are $i=1,...,M$ subjects and that each subject is measured n_i times. The Laird and Ware (1982) model is:

$$\underset{n_i\times 1}{\mathbf{y}_i} = \underset{n_i\times p}{\mathbf{X}_i}\,\underset{p\times 1}{\boldsymbol{\alpha}} + \underset{n_i\times k}{\mathbf{Z}_i}\,\underset{k\times 1}{\mathbf{b}_i} + \underset{n_i\times 1}{\mathbf{e}_i} \tag{2.7}$$

where

$$\mathbf{e}_i \sim N(\mathbf{0}, \underset{n_i\times n_i}{\mathbf{R}_i}) \tag{2.8}$$

and

$$\mathbf{b}_i \sim N(\mathbf{0}, \underset{k\times k}{\mathbf{D}}) \,. \tag{2.9}$$

In the above model $\mathbf{y}_i$ is the vector of observations for the ith individual. $\mathbf{X}_i$ and $\mathbf{Z}_i$ are known design matrices. $\boldsymbol{\alpha}$ is an unknown vector of fixed effects that needs to be estimated. The $\mathbf{b}_i$ is a random vector of unknown individual effects and $\mathbf{e}_i$ is the random error. In addition it is assumed that $\mathbf{e}_i$ and $\mathbf{b}_i$ are independent. It is usually assumed that $\mathbf{R}_i = \sigma^2\mathbf{I}_{n_i}$. The elements of $\mathbf{R}_i$ and $\mathbf{D}$ make up what is commonly known as the "variance components." Estimation can be carried out via the EM algorithm. The EM

algorithm alternates between calculating the conditional expected values and then maximizing the likelihood. There are several advantages to using the EM algorithm. First, for most practical purposes it is guaranteed to converge. Second, the variance components remain in the parameter space. Third, the EM algorithm produces conditional estimates of the individual random effects (Dempster, Laird, and Rubin, 1977; Wu, 1983). Laird and Ware (1982) also discuss how the EM algorithm can be implemented to obtain restricted maximum likelihood estimates (RML). The idea behind RML is to maximize the part of the likelihood that is invariant to the fixed effects (Thompson, 1962; Patterson and Thompson, 1971). Jennrich and Schluchter (1986) considered the following model for analyzing unbalanced longitudinal data:

$$\underset{n_i\times 1}{\mathbf{y}_i} = \underset{n_i\times p}{\mathbf{X}_i}\,\underset{p\times 1}{\boldsymbol{\beta}} + \underset{n_i\times 1}{\mathbf{e}_i}\,. \tag{2.10}$$

The $\mathbf{y}_i$ is the data vector for the ith individual (i=1,2,...,M). $\mathbf{X}_i$ is a known design matrix. $\boldsymbol{\beta}$ is a vector of unknown parameters to be estimated. The $\mathbf{e}_i$ is a vector of random errors distributed as $N(\mathbf{0}, \underset{n_i\times n_i}{\boldsymbol{\Sigma}_i})$. The $\boldsymbol{\Sigma}_i$ consists of q covariance parameters contained in the vector $\boldsymbol{\theta}$. Jennrich and Schluchter (1986) discuss several variance structures. The simplest variance structure is the completely independent covariance structure:

$$\boldsymbol{\Sigma}_i = \sigma^2\mathbf{I}_{n_i}\,. \tag{2.11}$$

The above variance structure can easily be modified to allow for different variances among groups of individuals. The next case they consider is the random-coefficients model:

$$\boldsymbol{\Sigma}_i = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i' + \sigma^2 \mathbf{I}_{n_i} . \tag{2.12}$$

This is similar to model (2.7) and is derived from model (2.10) by letting:

$$\underset{n_i \times 1}{\mathbf{e}_i} = \underset{n_i \times k}{\mathbf{Z}_i} \underset{k \times 1}{\mathbf{b}_i} + \underset{n_i \times 1}{\mathbf{u}_i} \text{ where } \mathbf{b}_i \sim N(\mathbf{0}, \underset{k \times k}{\mathbf{D}}) \text{ and independently } \mathbf{u}_i \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_{n_i}) . \tag{2.13}$$

The special case where $\mathbf{Z}_i$ is a column vector of all ones leads to the well known between-within mixed ANOVA model called compound symmetric or uniform. The next error structure they consider is the first-order autoregressive or AR(1):

$$\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma} = \left[\sigma_{st} \right] \text{ where } \sigma_{st} = \sigma^2 \rho^{|s-t|} . \tag{2.14}$$

Related to the AR(1) error structure is the banded error structure:

$$\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma} = \left[\sigma_{st} \right] \text{ where } \sigma_{st} = \theta_r \text{ , } r = |s-t| + 1 . \tag{2.15}$$

The next to last error pattern they consider is unstructured or arbitrary:

$$\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n_i} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n_i} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n_i 1} & \sigma_{n_i 2} & \cdots & \sigma_{n_i n_i} \end{bmatrix} . \tag{2.16}$$

They also present but do not use a factor-analytic covariance structure. The log-likelihood for model (2.10) is:

$$\ln L = \text{Constant} - \frac{1}{2}\sum_{i=1}^{M} \log|\boldsymbol{\Sigma}_i| - \frac{1}{2}\sum_{i=1}^{M}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})'\boldsymbol{\Sigma}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}). \qquad (2.17)$$

Jennrich and Schluchter (1986) discuss how to maximize (2.17) using Newton-Raphson, Fisher scoring, and a modified EM algorithm. Hocking (1985) discusses a model similar to that of model (2.10). His model is:

$$\underset{n\times 1}{\mathbf{Y}} = \underset{n\times p}{\mathbf{X}}\,\underset{p\times 1}{\boldsymbol{\alpha}} + \underset{n\times 1}{\mathbf{e}} \text{ where } \mathbf{e} \sim N(\mathbf{0}, \mathbf{V}) \text{ and } \underset{n\times n}{\mathbf{V}} = \sum_q \theta_q \mathbf{V}_q. \qquad (2.18)$$

Unlike Jennrich and Schluchter (1986), Hocking (1985) does not require that individuals be independent. Hocking (1985) gives an algorithm for finding maximum as well as restricted maximum likelihood estimates for model (2.18). Zerbe, Wu, and Zucker (1994) show that the usual Laird-Ware (1982) model is a special case of the Hocking (1985) model. All error structures discussed by Jennrich and Schluchter (1986) with the exception of AR(1) and factor-analytic can be handled by Hocking's (1985) model. Jones (1990, 1991, 1993) extends model (2.7) to allow for an AR(1) within subjects error structure along with random effects. Jones (1993) shows how the Laird-Ware model can be put into state space form and then how the Kalman filter can be used to calculate the likelihood. He uses this approach to find the maximum likelihood estimates.

2.4 A Brief History of the Mixed Model

As stated previously, Laird and Ware (1982) popularized the use of the mixed effects model to analyze longitudinal data. Biostatisticians tend to call model (2.7) the "Laird-Ware" model as if Laird and Ware had invented it. In fact the history of the mixed model is extensive and dates back to Airy (1861), although he didn't call it a mixed model, he used a mixed model approach to model astronomical observation. Fisher (1925) is credited with developing the first method of estimating variance components. While trying to estimate the intraclass correlation coefficient he set the ANOVA sums of squares equal to their expectation and solved a system of linear equations to obtain estimates for the variance components. This method would come to be known as the "method of moments technique." From the late 1930's to the late 1960's many authors used the "method of moments technique" to arrive at estimates of the variance components for several special cases of model (2.7). Henderson (1953) published a landmark paper for dealing with model (2.7). In his paper he gave three methods for estimating the variance components of unbalanced mixed models. His methods could be applied to as many crossed and/or nested classifications as necessary. A problem with the "method of moments technique" is that it can lead to negative estimates for variance components. A few authors have found closed form maximum likelihood estimates for model (2.7), but these only applied to simple cases of model (2.7). Hartley and Rao (1967) showed how the steepest ascent method could be used to obtain maximum likelihood estimates for a special case of model (2.7) where $\mathbf{D}$ is diagonal and $\mathbf{R} = \sigma^2\mathbf{I}$. With computers becoming more and more powerful maximum

likelihood methods became more and more popular. Finally, Harville (1976, 1977) generalized the Gauss-Markov theorem to the mixed effects model (2.7). His work became the impetus of Laird and Ware's contribution to the analysis of longitudinal data. Recently (1994), SAS released PROC MIXED for the personal computer. PROC MIXED implements model (2.7) and will no doubt become the "work horse" for researchers modeling linear longitudinal data. Searle (1992) has devoted a whole book to the mixed model. In his book he provides a whole chapter on the history of the mixed effects model. Much of what has previously been written in this thesis was taken from his book. All of the previously mentioned models have been geared for analyzing a single response variable, the next section will discuss models for multiple responses.

2.5 A Multi-response Linear Mixed Effects Model

Zucker, Zerbe, and Wu (in press) discuss a linear multi-response model. They have two pulmonary function outcomes of interest. First, they measure forced expiratory volume in one second (FEV_1). Second, they measure functional residual capacity (FRC). These measurements are repeated at approximately 3 month intervals for about 3 years. They fit lines to the FEV_1 and FRC data over time. Using a multi-response model they are able to answer questions like: Is the slope of FEV_1 related to the intercept of FRC? Their subject-time specific model is equivalent to:

$$\underset{n_{ij}\times 1}{\mathbf{y}_{ij}} = \underset{n_{ij}\times p\ p\times 1}{\mathbf{X}_{ij}\ \boldsymbol{\alpha}} + \underset{n_{ij}\times k\ k\times 1}{\mathbf{Z}_{ij}\ \mathbf{b}_i} + \underset{n_{ij}\times r\ r\times 1}{\mathbf{U}_{ij}\ \mathbf{e}_{ij}}. \qquad (2.19)$$

The $\mathbf{y}_{ij}$ is a vector of the multiple responses for the ith subject at the jth measurement time (i=1,...,M and j=1,...,m_i). The $\boldsymbol{\alpha}$ is a vector of fixed effects. The $\mathbf{b}_i$ is a vector of subject specific random effects. The $\mathbf{e}_{ij}$ are subject-time specific random errors. The $\mathbf{X}_{ij}$, $\mathbf{Z}_{ij}$, and $\mathbf{U}_{ij}$ are known design matrices. It is further assumed that:

$$\mathbf{b}_i \sim N(\mathbf{0}, \underset{k\times k}{\mathbf{D}}) \text{ and independently } \mathbf{e}_{ij} \sim N(\mathbf{0}, \underset{r\times r}{\mathbf{S}}). \tag{2.20}$$

Letting,

$$\mathbf{y}_i = \begin{bmatrix} \mathbf{y}_{i1} \\ \vdots \\ \mathbf{y}_{im_i} \end{bmatrix}, \mathbf{X}_i = \begin{bmatrix} \mathbf{X}_{i1} \\ \vdots \\ \mathbf{X}_{im_i} \end{bmatrix}, \mathbf{Z}_i = \begin{bmatrix} \mathbf{Z}_{i1} \\ \vdots \\ \mathbf{Z}_{im_i} \end{bmatrix}, \mathbf{e}_i = \begin{bmatrix} \mathbf{e}_{i1} \\ \vdots \\ \mathbf{e}_{im_i} \end{bmatrix} \text{ and } \mathbf{U}_i = \begin{bmatrix} \mathbf{U}_{i1} & & \\ & \ddots & \\ & & \mathbf{U}_{im_i} \end{bmatrix}$$

the following subject specific model is obtained:

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\alpha} + \mathbf{Z}_i\mathbf{b}_i + \mathbf{U}_i\mathbf{e}_i. \tag{2.21}$$

With the exception of the $\mathbf{U}_i$ this is essentially the Laird-Ware model (2.7). The $\mathbf{U}_i$ is an indicator matrix used to aid in the handling of missing data. For example, if bivariate data such as FEV_1 and FRC $(y_{FEV}, y_{FRC})'$ are collected on individuals over time, and at one of the times the FEV_1 value was missing then:

$$\mathbf{U}_{ij} = \begin{bmatrix} 0 & 1 \end{bmatrix}$$

for that subject-time specific model. The form of the residual error covariance matrix is also slightly different from what is normally used in the Laird-Ware model. In model (2.21) it can be shown that:

$$\mathbf{e}_i \sim N(\mathbf{0}, \mathbf{V} = \mathbf{I}_{m_i} \otimes \underset{\sim}{\mathbf{S}}),$$

where $\otimes$ is the Kronecker product. In the Laird-Ware (1982) model it is usually assumed that $\mathbf{e}_i \sim N(\mathbf{0}, \sigma^2\mathbf{I})$. Estimation can be carried out using simple modifications to the EM algorithm discussed by Laird and Ware (1982). Recently, Mickey et al. (1994) published a multi-response model similar to (2.21). However, their model is not as general. It requires every response to be measured at each time point.

Jones (1993) devotes a whole chapter to the fitting of multi-response models with random effects. His approach involves setting the problem up in state space form and using the Kalman filter to calculate the likelihood.

A natural extension of the linear mixed effects model seems to be the nonlinear mixed effects model. These models will be discussed next.

2.6 Nonlinear Mixed Effects Models

In many applications of longitudinal data, nonlinear models are preferable to the polynomial models fit using linear mixed effects models. Higher order polynomials can usually be fit to any set of data, however, these types of models are <u>not</u> based on any underlying biological theory. They also have difficulty fitting data that have asymptotes. Many nonlinear models have parameters that have biological interpretation and the curve

is derived from biological plausibility. Historically, there have been two approaches to modeling nonlinear repeated measures data. The first method is to just ignore the within subject correlation and fit a nonlinear model to the data via least squares. The second method is to use a two stage approach where each subject is first fit with a nonlinear least squares model. In the second stage, the parameters for each subject are used as the dependent variables in a Multivariate Analysis of Variance (MANOVA) to obtain estimates of the population parameters (Sheiner and Beal, 1980). In recent years the methods have become more sophisticated and warrant closer examination.

2.7 Nonlinear Longitudinal Data Analysis via Empirical Baye's Estimates

Berkey (1982) discussed the following nonlinear model for analyzing growth curves:

$$\underset{n_i\times 1}{\mathbf{y}_i} = \underset{q\times 1}{\mathbf{f}}(\underset{n_i\times 1}{\mathbf{b}_i}, \underset{n_i\times 1}{\mathbf{x}_i}) + \underset{n_i\times 1}{\mathbf{e}_i}. \tag{2.22}$$

The vector $\mathbf{y}_i$ is the data for the ith subject (i=1,...,M). The $\mathbf{f}$ is a vector valued function. The $\mathbf{b}_i$ are subject specific parameters that need to be estimated. The $\mathbf{x}_i$ are covariates associated with the ith subject. The $\mathbf{e}_i$ are random errors. The following is assumed in the above model:

$$\mathbf{e}_i|\mathbf{b}_i \sim N(\mathbf{0}, \sigma_i^2\mathbf{I}). \tag{2.23}$$

Based on the above assumption the conditional distribution function of $\mathbf{y}_i$ is:

$$g(\mathbf{y}_i|\mathbf{b}_i, \sigma_i^2, \mathbf{x}_i) = 1/[(2\pi\sigma_i^2)^{n_i/2}] \exp\left\{\frac{-1}{2\sigma_i^2}\sum_{j=1}^{n_i}[y_{ij} - f(\mathbf{b}_i, x_{ij})]^2\right\}. \tag{2.24}$$

The curve for an individual is completely determined by $\mathbf{b}_i$. This curve can be considered to be a realization of a random variable **B** having probability density function $h_{\mathbf{B}}(\mathbf{b}_i)$ that is multivariate normal with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. Therefore the density function of $\mathbf{b}_i$ is:

$$h(\mathbf{b}_i|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{q/2}\sqrt{|\boldsymbol{\Sigma}|}} \exp\{\tfrac{-1}{2}(\mathbf{b}_i - \boldsymbol{\mu})' \underset{q\times q}{\boldsymbol{\Sigma}}^{-1}(\mathbf{b}_i - \boldsymbol{\mu})\}. \tag{2.25}$$

It is further assumed that $\mathbf{b}_i$ and σ^2 are independent, this implies that the joint distribution is:

$$q(\mathbf{b}_i, \sigma^2) = h(\mathbf{b}_i)p(\sigma^2). \tag{2.26}$$

The noninformative prior density $p(\sigma^2) \propto 1/\sigma^2$ will be used. Following the work of Lindley and Smith (1972), point estimates for $\mathbf{b}_i$ will be sought by finding the mode (not mean) of the posterior distribution. The posterior distribution can be shown to be:

$$w(\mathbf{b}_i, \sigma^2|\mathbf{y}_i) \propto \frac{1}{\sigma^{n_i+2}} \exp\left\{\tfrac{-1}{2}\left[[\mathbf{y}_i - \mathbf{f}(\mathbf{b}_i, \mathbf{x}_i)]'[\mathbf{y}_i - \mathbf{f}(\mathbf{b}_i, \mathbf{x}_i)]/\sigma^2\right] + (\mathbf{b}_i - \boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\mathbf{b}_i - \boldsymbol{\mu})\right\}. \tag{2.27}$$

Estimation can be carried out by searching for the joint mode of the posterior distribution or by integrating out σ^2 in (2.27) to obtain the marginal distribution $k(\mathbf{b}_i|\mathbf{y}_i)$. Finding the $\mathbf{b}_i$ that maximizes $k(\mathbf{b}_i|\mathbf{y}_i)$ is the same as finding the $\mathbf{b}_i$ that minimizes $1/k(\mathbf{b}_i|\mathbf{y}_i)$. The marginal posterior mode is the $\mathbf{b}_i$ that minimizes:

$$1/k(\mathbf{b}_i|\mathbf{y}_i) \propto [\exp\{\tfrac{1}{2}(\mathbf{b}_i-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{b}_i-\boldsymbol{\mu})\}][\{\mathbf{y}_i-\mathbf{f}(\mathbf{b}_i,\mathbf{x}_i)\}'\{\mathbf{y}_i-\mathbf{f}(\mathbf{b}_i,\mathbf{x}_i)\}]^{\frac{1}{2}n_i}. \qquad (2.28)$$

Notice that as n_i increases, less weight is given to the prior. Berkey (1982) is interested in improving subject specific parameter estimates by borrowing information from other subjects to estimate the parameters for a particular individual. This is a variant of Stein shrinkage (James and Stein, 1961). Berkey (1982) is not interested in the population parameter estimates. Racine-Poon (1985) extends the work of Berkey (1982) by adding a 3rd stage to the hierarchical model. Adding the third stage allows for the estimation of population parameters. Where Berkey (1982) considers the prior information on $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ known (or at least a good estimate is known) Racine-Poon (1985) considers them to arise from a vague prior distribution. In doing so, Racine-Poon (1985) is able to obtain empirical Bayes estimates of population parameters as well as individual parameters.

Another approach for analyzing longitudinal data has been motivated by the generalized linear model (GLIM). This approach will be discussed next.

2.8 A Generalized Estimating Equations Approach

Liang and Zeger (1986) extend the generalized linear model (GLIM) popularized by McCullagh and Nelder (1983) to the case of longitudinal data. Under mild assumptions their estimating equations give consistent estimates of the regression parameters and their variances. In the discussion that follows we will use the following notation. Let y_{ij} be the jth observation on the ith subject (j=1,...,n_i and i=1,...,M). Let $\mathbf{x}_{ij}$ be a $p \times 1$ vector of known covariates associated with y_{ij}. McCullagh and Nelder (1983) discuss the case where n_i =1 for all subjects. For discussing this case the j can be dropped from the notation. In the quasi-likelihood approach discussed by McCullagh and Nelder (1983) the mean is related to the explanatory variables via a 'link' function:

$$g(\mu_i) = \underset{1\times p}{\mathbf{x}_i'}\underset{p\times 1}{\boldsymbol{\beta}}. \tag{2.29}$$

And the variance is related to the mean through a variance function:

$$var(y_i) = v(\mu_i)\phi = v_i. \tag{2.30}$$

The ϕ is a scale parameter and interest usually lies in estimating the parameter vector $\boldsymbol{\beta}$. This can be accomplished by solving the generalized estimating equations:

$$U_k(\boldsymbol{\beta}) = \sum_{i=1}^{M}(\partial\mu_i/\partial\beta_k)v_i^{-1}(y_i - \mu_i) = 0. \tag{2.31}$$

The solutions can be obtained via iteratively reweighted least squares. So far the y_i have all been independent. However, for longitudinal data the observations on a subject are correlated. If $\mathbf{y}_i$ is a vector of observations, Liang and Zeger (1986) handle the within subject correlation by defining a $n_i \times n_i$ "working" correlation matrix $\underset{s\times 1}{\mathbf{R}_i(\boldsymbol{\alpha})}$ for the ith subject. Different structures for the "working" correlation matrix will be discussed later. Based on the "working" correlation matrix the following "working" variance-covariance matrix can be defined:

$$\underset{n_i\times n_i}{\mathbf{V}_i} = \mathbf{A}_i^{1/2}\mathbf{R}_i(\boldsymbol{\alpha})\mathbf{A}_i^{1/2}\phi \tag{2.32}$$

where

$$\mathbf{A}_i = diag\{v(\mu_{i1}), ..., v(\mu_{in_i})\}. \tag{2.33}$$

Note that if the "working" correlation matrix were the "true" correlation matrix then $\mathbf{V}_i$ would be the "true" variance-covariance matrix of $\mathbf{y}_i$. Liang and Zeger (1986) extend (2.31) as follows:

$$\sum_{i=1}^{M}\mathbf{U}_i(\boldsymbol{\alpha},\boldsymbol{\beta}) = \sum_{i=1}^{M}(\partial\boldsymbol{\mu}_i/\partial\boldsymbol{\beta})'\mathbf{V}_i^{-1}(\mathbf{y}_i-\boldsymbol{\mu}_i) = \mathbf{0} \tag{2.34}$$

where $\mathbf{y}_i = (y_{i1}, y_{i2}, ..., y_{in_i})'$ and $\boldsymbol{\mu}_i = (\mu_{i1}, \mu_{i2}, ..., \mu_{in_i})'$. Given an estimate of $\mathbf{R}_i(\boldsymbol{\alpha})$ and ϕ, an updated estimate of $\boldsymbol{\beta}$ can be obtained by iteratively reweighted least squares. Next, given estimates of $\boldsymbol{\beta}$, standardized residuals can be calculated and then used to get new consistent estimates of $\mathbf{R}_i(\boldsymbol{\alpha})$ and ϕ. This process keeps repeating until convergence. There are several choices for selecting the form of the "working" correlation matrix. The possibilities include:

1. The simplest case is to let $\mathbf{R}_i(\boldsymbol{\alpha}) = \mathbf{I}_{n_i}$, then the model essentially reduces to the case discussed by McCullagh and Nelder (1983).

2. If all subjects are measured at the same times/doses and there are no missing data then the correlation matrix can be totally unstructured. If n_i=n for all subjects then there will be $n(n\text{-}1)/2$ correlations.

3. Another possibility is to let $[\mathbf{R}_i(\alpha)]_{jk} = \alpha \quad j \neq k$, this is the correlation structure assumed in the compound symmetric random effects model.

4. Another choice is:

$$[\mathbf{R}_i(\alpha)]_{jk} = \begin{Bmatrix} \alpha^{|t_{ij}-t_{ik}|} & |t_{ij}-t_{ik}| \leq m \\ 0 & |t_{ij}-t_{ik}| > m \end{Bmatrix}. \tag{2.35}$$

The t_{ij} and t_{ik} are the jth and kth observation times for the ith subject. This correlation structure is known as a stationary m-dependent process.

Zeger, Liang, and Albert (1988) introduce terminology that has become popular when discussing longitudinal models. They classify models as either Population Average (PA) or Subject Specific (SS).

In the PA model interest lies in the marginal expectation, $\mu_{it} = E(y_{it})$. We assume that the link function and variance function are as shown in (2.29) and (2.30). For example we could have:

$$\text{logit}(\mu_{it}) = \mathbf{x}_{it}'\boldsymbol{\beta}^{PA} \text{ and } Var(y_{it}) = \mu_{it}(1-\mu_{it}).$$

The estimate of $\boldsymbol{\beta}$ describes how the population averaged response depends on the covariates.

In the Subject Specific (SS) model the conditional mean $u_{it} = E(y_{it}|\mathbf{b}_i)$ is the focus. The link and variance functions are:

$$g_{ss}(u_{it}) = \mathbf{x}_{it}'\boldsymbol{\beta}^{ss} + \underset{1\times q}{\mathbf{z}_i'}\underset{q\times 1}{\mathbf{b}_i} \text{ and } Var(y_{it}|\mathbf{b}_i) = \mathbf{V}_{ss}(u_{it})\phi, \tag{2.36}$$

where $\mathbf{b}_i$ is a random effect from a mixture distribution F. For example the following formulation could be used:

$$\text{logit}(u_{it}) = \mathbf{x}_{it}'\boldsymbol{\beta}^{ss} + \mathbf{z}_i'\mathbf{b}_i, \tag{2.37}$$

$$Var(y_{it}|\mathbf{b}_i) = u_{it}(1-u_{it}), \tag{2.38}$$

$$\mathbf{b}_i \sim N(\mathbf{0}, \underset{q\times q}{\mathbf{D}}). \tag{2.39}$$

Since $E(\mathbf{b}_i) = \mathbf{0}$, $\boldsymbol{\beta}^{ss}$ describes on average how an individual response depends on the covariates. Zeger et al. (1988) summarize the distinction between population average (PA) models and subject specific (SS) models this way:

> The principle distinction between SS and PA models is whether the regression coefficients describe an individual's or the average population response to changing **x**. A secondary distinction is in the nature of the assumed time dependence. PA models only describe the covariance among repeated observations for a subject; SS models explain the source of this covariance. In PA models the covariance matrix must be positive-definite but is otherwise unrestricted. In SS models, the time dependence arises solely from the shared subject effects, $\mathbf{b}_i$, in the conditional mean. The covariance matrix is thus fully determined by the choices of $g(u_{it})$ and F.

Zeger et al. (1988) point out that in the linear model the fixed effects for PA and SS models have the same interpretation.

The strength of Zeger and Liang's (1986) quasi-likelihood approach is that it can handle non-Guassian situations. Traditionally, the generalized estimating equations (GEE) approach has focused on a few well known link functions such as the identity (linear), logit, probit, and log. As pointed out by Lindstrom and Bates (1990), the GLIM SS model is restricted by the fact that the link function in (2.36) is a function of one variable and is therefore more restrictive than need be. The GLIM model does not suffer from a major drawback of Berkey (1982) and Racine-Poon (1985), in that the method does not require that each individual must be fit with a nonlinear model. Another drawback to the methods of Berkey (1982) and Racine-Poon (1985) is that all parameters are considered to vary stochastically across individuals. Other miscellaneous work on nonlinear mixed effects model will be discussed next.

2.9 Other Work on Nonlinear Mixed Effects Models

Jones (1993) devotes a whole chapter to the fitting of nonlinear models with random effects. Again his approach involves casting the problem into state space form and using the Kalman filter to evaluate the likelihood.

All of the nonlinear models discussed so far have been geared toward single response data. Seber and Wild (1989) discusses models of the following form:

$$y_{ij} = f_j(\mathbf{X}_i, \boldsymbol{\beta}) + \varepsilon_{ij} \qquad (i = 1, 2, ..., n, \quad j = 1, 2, ..., d). \tag{2.40}$$

Seber and Wild (1989) call this a "multi-response" model. In this model, there are d nonlinear models, each observed at n values of $\mathbf{X}$. Model (2.40) can also be put into vector notation, the model becomes:

$$\underset{d\times 1}{\mathbf{y}_i} = \underset{d\times 1}{\mathbf{f}(\mathbf{X}_i, \boldsymbol{\beta})} + \underset{d\times 1}{\boldsymbol{\varepsilon}_i}, \tag{2.41}$$

where

$$\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \underset{d\times d}{\boldsymbol{\Sigma}}), \tag{2.42}$$

and

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1d} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{d1} & \sigma_{d2} & \cdots & \sigma_{dd} \end{bmatrix}. \tag{2.43}$$

The original intent of this model was to handle multiple response data. However the error structure used would be the multivariate analog of $\sigma^2\mathbf{I}$. The model can be used to analyze longitudinal data if we think of the d responses as being d measurements on the same subject at d different times and if we drop the j subscript for the nonlinear functions so that we have only one function. Then model (2.40) can be used to analyze single response nonlinear longitudinal data. The resulting covariance structure will be arbitrary. The -2 log likelihood for this model is:

$$-2\ln L = nd\ln(2\pi) + n\ln|\boldsymbol{\Sigma}| + \sum_{i=1}^{n}[\mathbf{y}_i - \mathbf{f}(\mathbf{X}_i, \boldsymbol{\beta})]' \boldsymbol{\Sigma}^{-1} [\mathbf{y}_i - \mathbf{f}(\mathbf{X}_i, \boldsymbol{\beta})]. \tag{2.44}$$

A method for minimizing (2.44) is given by Bates and Watts (1987).

Magnus and Neudecker (1988) describe a very general nonlinear mixed effects model. Their model is:

$$\underset{n\times 1}{\mathbf{y}} = \boldsymbol{\mu}(\underset{p\times 1}{\boldsymbol{\alpha}}, \underset{r\times 1}{\boldsymbol{\tau}}) + \underset{n\times 1}{\mathbf{e}} \text{ where } \mathbf{e} \sim N(\mathbf{0}, \mathbf{V}[\underset{q\times 1}{\boldsymbol{\theta}}, \boldsymbol{\tau}]). \tag{2.45}$$

The $\mathbf{y}$ is a vector of observations. The $\boldsymbol{\mu}$ is a nonlinear vector valued function that depends on the parameter vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\tau}$. The error is normally distributed with mean $\mathbf{0}$ and variance $\mathbf{V}$ that depends on the parameter vectors $\boldsymbol{\theta}$ and $\boldsymbol{\tau}$. Notice that both the expectation and variance depend on the same parameter $\boldsymbol{\tau}$. The log likelihood for model (2.45) is:

$$\ln L = -(n/2)\ln(2\pi) - (\ln|\mathbf{V}|)/2 - (\mathbf{y}-\boldsymbol{\mu})'\mathbf{V}^{-1}(\mathbf{y}-\boldsymbol{\mu})/2 \ . \tag{2.46}$$

Maximization of (2.46) can be carried out by taking derivatives of (2.46) with respect to $\boldsymbol{\alpha}$, $\boldsymbol{\tau}$ and $\boldsymbol{\theta}$ and equating them to zero. Doing so results in the following "normal" equations:

$$(\partial\boldsymbol{\mu}/\partial\boldsymbol{\alpha}_h)'\mathbf{V}^{-1}(\mathbf{y}-\boldsymbol{\mu}) = 0 \qquad (h = 1,...,p) \tag{2.47}$$

$$trace[\mathbf{V}^{-1}(\partial\mathbf{V}/\partial\theta_t)] - (\mathbf{y}-\boldsymbol{\mu})'\mathbf{V}^{-1}(\partial\mathbf{V}/\partial\theta_t)\mathbf{V}^{-1}(\mathbf{y}-\boldsymbol{\mu}) = 0 \qquad (t = 1,...,q) \tag{2.48}$$

$$trace[\mathbf{V}^{-1}(\partial\mathbf{V}/\partial\tau_k)] - (\mathbf{y}-\boldsymbol{\mu})'\mathbf{V}^{-1}(\partial\mathbf{V}/\partial\tau_k)\mathbf{V}^{-1}(\mathbf{y}-\boldsymbol{\mu}) - 2(\partial\boldsymbol{\mu}/\partial\tau_k)\mathbf{V}^{-1}(\mathbf{y}-\boldsymbol{\mu}) = 0 \ . \tag{2.49}$$
$$(k = 1,...,r)$$

The solution to the above system of equations will yield the maximum likelihood estimates. Typically, except for some special cases of model (2.45) these solutions are extremely difficult to obtain. Magnus and Neudecker (1988) offer no methods for estimation in model (2.45). Also, the model is not written in a subject specific form. This makes implementation of the model difficult or even impossible if n is large. Model (2.45) is good from a theoretical point of view. Once a model can be shown to be a special case of model (2.45) all of the theoretical properties derived by Magnus and Nuedecker (1988) can be applied.

2.10 Nonlinear Mixed Effects Models Directly Relating to MNLMEM

The multi-response nonlinear mixed effects model (MNLMEM) is essentially an extension of the work by Sheiner and Beal (1980), Lindstrom and Bates (1990), Hirst et al. (1991), and Vonesh and Carter (1992). A discussion of the strengths and weaknesses of the various methods will prove useful in our discussion of MNLMEM. In order to facilitate the comparison of several different methods a single model will be used. As a result some of the details of the various methods will be omitted. The basic model that the previous authors wish to make inference about can be written as:

$$\underset{n_i\times 1}{\mathbf{y}_i}\,|\mathbf{b}_i \sim N[\mathbf{f}(\underset{p\times 1}{\boldsymbol{\alpha}}, \underset{q\times 1}{\mathbf{b}_i}, \underset{k_i\times 1}{\mathbf{x}_i}), \sigma^2\mathbf{I}_{n_i}] \tag{2.50}$$

and

$$\mathbf{b}_i \sim N[\mathbf{0}, \mathbf{D}]. \tag{2.51}$$

The vector $\mathbf{y}_i$ contains the data for the ith subject (i=1,...,M). The $\mathbf{f}$ is a nonlinear vector valued function. The $\boldsymbol{\alpha}$ is an unknown vector of fixed effects and the $\mathbf{b}_i$ is a vector of random effects. The $\mathbf{x}_i$ is a vector of covariables associated with the ith individual, this is usually time or dose. Some authors don't assume normality in (2.50) and (2.51). The problem with the nonlinear mixed effects model (2.50 and 2.51) is that traditional methods of estimation, like maximum likelihood, can not be performed directly on the model because the likelihood surface for the model is very complicated. For example, in order to find the likelihood, the probability density function (pdf) for $\mathbf{y}_i$ needs to be found. The pdf is given by:

$$g(\mathbf{y}_i) = \int g(\mathbf{y}_i|\mathbf{b}_i)g(\mathbf{b}_i)d\mathbf{b}_i$$

$$= \int (2\pi\sigma^2)^{-n_i/2} \exp\{\tfrac{-1}{2\sigma^2}[\mathbf{y}_i - \mathbf{f}(\boldsymbol{\alpha}, \mathbf{b}_i, \mathbf{x}_i)]'[\mathbf{y}_i - \mathbf{f}(\boldsymbol{\alpha}, \mathbf{b}_i, \mathbf{x}_i)]\}(2\pi)^{-q/2}|\mathbf{D}|^{-1/2} \exp\{\tfrac{-1}{2}\mathbf{b}_i'\mathbf{D}^{-1}\mathbf{b}_i\}d\mathbf{b}_i. \quad (2.52)$$

There is no closed form solution for this integral because we are integrating with respect to $\mathbf{b}_i$ and $\mathbf{b}_i$ is an argument of a nonlinear function. Thus, except for a few special cases, it is impossible to obtain the likelihood function for the nonlinear mixed effects model. However, it may be possible to approximate the likelihood surface near the values that maximize the original likelihood. If the contours for the approximate surface closely match the contours for the original surface near the estimates, then inferences made on the approximate model will also apply to the original model. All of the previously mentioned authors deal with this problem by using various Taylor series approximations to the model. After the model has been "linearized" via a first order Taylor series

expansion, various methods of estimation are employed on the approximated model. The hope is that the approximation is good and results can be inferred back to the original model. How well the Taylor series approximates the original model is a problem that will briefly be discussed in Chapter V.

Sheiner and Beal (1980) approximate model (2.50) by expanding about the expected value of the random effects, namely $E(\mathbf{b}_i)=\mathbf{0}$. Doing so results in the following approximate model:

$$\mathbf{y}_i|\mathbf{b}_i \overset{\bullet}{\sim} \left([\mathbf{f}(\boldsymbol{\alpha}, \mathbf{b}_i = \mathbf{0}, \mathbf{x}_i) + \mathbf{Z}_i(\boldsymbol{\alpha})\mathbf{b}_i], \sigma^2\mathbf{I}_{n_i}\right) \tag{2.53}$$

where

$$\mathbf{Z}_i(\boldsymbol{\alpha}) = \partial\mathbf{f}/\partial\mathbf{b}_i'|_{b_i=0}. \tag{2.54}$$

It is known that for any two random variables X and Y the following holds:

$$E(X) = E[E(X|Y)] \text{ and } Var(X) = E[Var(X|Y)] + Var[E(X|Y)].$$

This fact coupled with (2.51) can be used to approximate the marginal distribution of $\mathbf{y}_i$. Thus the approximate model is:

$$\mathbf{y}_i \overset{\bullet}{\sim} \left(\mathbf{f}(\boldsymbol{\alpha}, \mathbf{b}_i = \mathbf{0}, \mathbf{x}_i), \{\mathbf{Z}_i(\boldsymbol{\alpha})\mathbf{D}\mathbf{Z}_i'(\boldsymbol{\alpha}) + \sigma^2\mathbf{I}_{n_i}\}\right). \tag{2.55}$$

The notation that has been used above is useful for describing the hierarchical distribution theory of the nonlinear mixed effects model. However, when describing models it is often easier to work with a more compact notation. We can arrive at the

same approximate model (2.55) using slightly different notation. The following notation will be now be adopted. Let the original model be written as:

$$\mathbf{y}_i = \mathbf{f}(\boldsymbol{\alpha}, \mathbf{b}_i, \mathbf{x}_i) + \mathbf{e}_i . \tag{2.56}$$

Sheiner and Beal (1980) expand (2.56) about the expected value of the random effects [$E(\mathbf{b}_i)=\mathbf{0}$]. The approximate model becomes:

$$\mathbf{y}_i \approx \mathbf{f}(\boldsymbol{\alpha}, \mathbf{b}_i = \mathbf{0}, \mathbf{x}_i) + \mathbf{Z}_i(\boldsymbol{\alpha})[\mathbf{b}_i - \mathbf{0}] + \mathbf{e}_i \tag{2.57}$$

where as before

$$\mathbf{Z}_i(\boldsymbol{\alpha}) = \partial \mathbf{f} / \partial \mathbf{b}_i' \big|_{b_i=0} .$$

If we make the following assumption:

$$\mathbf{b}_i \sim (\mathbf{0}, \mathbf{D}) \text{ and independently } \mathbf{e}_i \sim (\mathbf{0}, \sigma^2 \mathbf{I}_{n_i}), \tag{2.58}$$

then the approximate model (2.57) is the same as the approximate model (2.55). In the approximate model it is important to note that the matrix of partial derivatives depends on $\boldsymbol{\alpha}$. Sheiner and Beal (1980) advocate using extended least squares (ELS) to accomplish estimation. ELS is carried out by minimizing the following objective function:

$$\begin{aligned} O(\boldsymbol{\alpha}, \mathbf{D}, \sigma^2) = & \sum_{i=1}^{M} \log \left| \mathbf{Z}_i(\boldsymbol{\alpha}) \mathbf{D} \mathbf{Z}_i'(\boldsymbol{\alpha}) + \sigma^2 \mathbf{I}_{n_i} \right| \\ & + \sum_{i=1}^{M} [\mathbf{y}_i - \mathbf{f}(\boldsymbol{\alpha}, \mathbf{b}_i = \mathbf{0}, \mathbf{x}_i)]' (\mathbf{Z}_i(\boldsymbol{\alpha}) \mathbf{D} \mathbf{Z}_i'(\boldsymbol{\alpha}) + \sigma^2 \mathbf{I}_{n_i})^{-1} [\mathbf{y}_i - \mathbf{f}(\boldsymbol{\alpha}, \mathbf{b}_i = \mathbf{0}, \mathbf{x}_i)] . \end{aligned} \tag{2.59}$$

If normality is assumed in (2.58) then minimizing the previous objective function will yield maximum likelihood estimates for the approximate model (2.57). Also, assuming normality the approximate model is a special case of a model discussed by Magnus and Nuedecker (1988). This can be seen by comparing (2.45) and (2.55). Notice that both the expectation and variance depend on $\boldsymbol{\alpha}$. As pointed out by van Houwelingen (1988) this method of estimation can often lead to inconsistent estimates, He writes;

> In a series of papers in the pharmacological literature, Sheiner and Beal and others advocate the extended least squares (ELS) methodology that combines the regression and variance model into a single objective function based on normal-theory maximum likelihood. The inadequacy of this method is folklore in the (mathematical) statistical literature.

Vonesh and Carter (1992) prefer to work with a special case of model (2.56) where the random effects enter the model in a linear fashion. However, they recognize the importance of model (2.56) and approximate the model similar to Sheiner and Beal (1980) by expanding around the expected value of the random effects. However, instead of letting the resulting matrix of partial derivatives ($\mathbf{Z}_i$) depend on $\boldsymbol{\alpha}$ they substitute in the consistent ordinary least squares (OLS) estimate $\hat{\boldsymbol{\alpha}}_{OLS}$. They also avert the inconsistency problem by using generalized least squares (GLS) instead of ELS. The GLS objective function they minimize is:

$$O(\boldsymbol{\alpha}|\mathbf{y}_i, \hat{\boldsymbol{\alpha}}_{OLS}, \hat{\mathbf{D}}, \hat{\sigma}^2) = \sum_{i=1}^{M}[\mathbf{y}_i - \mathbf{f}(\boldsymbol{\alpha}, \mathbf{x}_i)]'[\mathbf{Z}_i(\hat{\boldsymbol{\alpha}}_{OLS})\hat{\mathbf{D}}\mathbf{Z}_i'(\hat{\boldsymbol{\alpha}}_{OLS}) + \hat{\sigma}^2\mathbf{I}_{n_i}]^{-1}[\mathbf{y}_i - \mathbf{f}(\boldsymbol{\alpha}, \mathbf{x}_i)]. \tag{2.60}$$

The estimates for the variance components $\boldsymbol{\theta} = \{\sigma^2, \mathbf{D}\}$ are arrived at via a method of moments technique. Vonesh and Carter (1992) argue that upon minimizing (2.60) one obtains a strongly consistent estimate of $\boldsymbol{\alpha}$ and they show that this estimate, $\hat{\boldsymbol{\alpha}}_{GLS}$, is asymptotically normally distributed with mean $\boldsymbol{\alpha}$ and variance given by:

$$\left[\sum_{i=1}^{M} \mathbf{Z}_i'(\hat{\boldsymbol{\alpha}}_{GLS})[\mathbf{Z}_i(\hat{\boldsymbol{\alpha}}_{OLS})\hat{\mathbf{D}}\mathbf{Z}_i'(\hat{\boldsymbol{\alpha}}_{OLS}) + \hat{\sigma}^2\mathbf{I}_{n_i}]^{-1}\mathbf{Z}_i(\hat{\boldsymbol{\alpha}}_{GLS})\right]^{-1}. \qquad (2.61)$$

A nice feature of the Vonesh and Carter (1992) method is that they do not require the assumption of normality. However, as they point out, their estimates are not asymptotically efficient. If the errors and random effects are indeed normally distributed and the variance structure is correctly specified then ELS (same as maximum likelihood) will beat Vonesh and Carter (1992). A problem arises in that small deviations in these assumptions can drastically reduce the efficiency. Another drawback with the Vonesh and Carter (1992) method is that the distribution theory for the variance components is lacking. Also, Vonesh and Carter (1992) require that each subject have a minimum amount of data and the amount of data required varies from model to model.

Hirst et al. (1991) combine the methods of Sheiner and Beal (1980) along with the traditional nonlinear techniques by expanding model (2.56) about current estimates of the fixed effects ($\boldsymbol{\alpha}_0$) and the expected value of the random effects ($\mathbf{b}_{i0}$=E[$\mathbf{b}_i$]=$\mathbf{0}$). Their approximate model becomes:

$$\mathbf{y}_i \approx \mathbf{f}(\boldsymbol{\alpha}_0, \mathbf{b}_{i0} = \mathbf{0}, \mathbf{x}_i) + \mathbf{H}_i(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0) + \mathbf{Z}_i(\mathbf{b}_i - \mathbf{0}) + \mathbf{e}_i \qquad (2.62)$$

where

$$\mathbf{H}_i = \partial \mathbf{f}/\partial \boldsymbol{\alpha}'\Big|_{\substack{\alpha=\alpha_0 \\ b_i=b_{i0}=0}} \quad \mathbf{Z}_i = \partial \mathbf{f}/\partial \mathbf{b}_i'\Big|_{\substack{\alpha=\alpha_0 \\ b_i=b_{i0}=0}} . \tag{2.63}$$

Unlike Sheiner and Beal (1980) these matrices of partial derivative do not contain any unknown parameters. The approximate model (2.62) can be written in the following form:

$$\mathbf{y}_i - \mathbf{f}(\boldsymbol{\alpha}_0, \mathbf{b}_{i0} = \mathbf{0}, \mathbf{x}_i) + \mathbf{H}_i\boldsymbol{\alpha}_0 \approx \mathbf{H}_i\boldsymbol{\alpha} + \mathbf{Z}_i\mathbf{b}_i + \mathbf{e}_i \tag{2.64}$$

or

$$\mathbf{y}_i^* \approx \mathbf{H}_i\boldsymbol{\alpha} + \mathbf{Z}_i\mathbf{b}_i + \mathbf{e}_i . \tag{2.65}$$

Assuming normality in (2.58), the objective function for the approximate model (2.65) is very similar to Sheiner and Beal's (2.59). The objective function is:

$$O(\boldsymbol{\alpha}, \mathbf{D}, \sigma^2 | \mathbf{y}_i) = \sum_{i=1}^{M} \log \left| \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i' + \sigma^2\mathbf{I}_{n_i} \right| + \sum_{i=1}^{M} [\mathbf{y}_i^* - \mathbf{H}_i\boldsymbol{\alpha}]' (\mathbf{Z}_i\mathbf{D}\mathbf{Z}_i' + \sigma^2\mathbf{I}_{n_i})^{-1} [\mathbf{y}_i^* - \mathbf{H}_i\boldsymbol{\alpha}] . \tag{2.66}$$

Minus a constant (2.66) is the -2 log likelihood for the approximate model (2.65). A big difference between the approximate model of Sheiner and Beal (1980) and the approximate model of Hirst et al. (1991) is that in the latter model the mean and the variance do not share an unknown parameter. According to van Houwelingen (1988), "Generally speaking, ELS is fine if the variance model does not depend on $\boldsymbol{\alpha}$". In terms of the approximate model, the $\mathbf{Z}_i$ can be considered fixed. The approximate model (2.65) can be recognized as a Laird-Ware (1982) model and maximum likelihood

estimation can be carried out using the EM algorithm. Once the EM algorithm converges, new estimates for the fixed effects ($\boldsymbol{\alpha}_1$), are available. These new estimates can then be used to update the approximate model (2.64) and the process is repeated until the likelihood of the approximate model (2.64) can not be improved. An advantage of using the EM algorithm in this scenario is that at the final iteration of the EM algorithm estimates of the random effects, $\mathbf{b}_i$, are readily available. Another advantage is that the EM algorithm can easily be used to obtain restricted maximum likelihood estimates (RML). Also, since the EM algorithm provides maximum likelihood estimates of the fixed effects and variance components we know that all estimates are asymptotically consistent and asymptotically efficient (Cassella and Berger, 1990). These results can be used to obtain estimates of the standard errors of the estimated fixed effects and variance components. This is an advantage over Vonesh and Carter (1992), since they only give the asymptotic properties for the fixed effects. It is important to keep in mind that the asymptotic properties that are being discussed are solely based on the linearized model. If the linearized model does not do a good job of approximating the original nonlinear model then the asymptotic properties discussed here may be very misleading. If we are willing to assume that the approximate model (2.64) closely approximates the real model (2.56) near the parameter estimates, we can infer the asymptotic theory of the approximate model to the original model. Another advantage of Hirst et al. (1991) is that there is no minimum amount of data needed for each individual. Hirst et al. (1991) also stress the special case of model (2.56) where the random effects enter the model in a linear fashion. This model can be represented as:

$$\mathbf{y}_i = \mathbf{f}(\boldsymbol{\alpha}, \mathbf{x}_i) + \mathbf{Z}_i\mathbf{b}_i + \mathbf{e}_i . \tag{2.67}$$

Assuming normality in (2.58), the exact likelihood of the above model (2.67) can be described. There is no need to approximate the likelihood surface. In order to find the maximum likelihood estimates, Hirst et al. (1991) expand about an initial estimate of the fixed effects ($\boldsymbol{\alpha}_0$) as a computational crutch. Their approach is similar to what is done in ordinary nonlinear regression. The model becomes:

$$\mathbf{y}_i \approx \mathbf{f}(\boldsymbol{\alpha}_0, \mathbf{x}_i) + \mathbf{H}_i(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0) + \mathbf{Z}_i\mathbf{b}_i + \mathbf{e}_i \tag{2.68}$$

where

$$\mathbf{H}_i = \partial \mathbf{f}/\partial \boldsymbol{\alpha}' \big|_{\alpha=\alpha_0} . \tag{2.69}$$

This model can be written as:

$$\mathbf{y}_i - \mathbf{f}(\boldsymbol{\alpha}_0, \mathbf{x}_i) + \mathbf{H}_i\boldsymbol{\alpha}_0 \approx \mathbf{H}_i\boldsymbol{\alpha} + \mathbf{Z}_i\mathbf{b}_i + \mathbf{e}_i \tag{2.70}$$

or

$$\mathbf{y}_i^* \approx \mathbf{H}_i\boldsymbol{\alpha} + \mathbf{Z}_i\mathbf{b}_i + \mathbf{e}_i . \tag{2.71}$$

The difference between this approximate model and the one given in (2.64) is that the $\mathbf{H}_i$ and $\mathbf{Z}_i$ are not matrices that depend on $\mathbf{b}_i$ since there was no expansion about $\mathbf{b}_i$. It should be recognized that model (2.71) is of the Laird-Ware (1982) form. Therefore, the EM algorithm can be used to find new maximum likelihood or restricted maximum likelihood estimates of the fixed effects. These new estimates can now be taken as the initial values in the approximate model (2.70). The process is repeated until the likelihood can be improved no more. At this point the EM algorithm has maximized the

likelihood for the original model (2.67) and estimates of the fixed effects, variance components, and random effects are available. Model (2.67) is very similar to the model that Vonesh and Carter (1992) prefer to work with. The difference between Vonesh and Carter (1992) and Hirst et al. (1991) concerning this model (2.67) is that Vonesh and Carter (1992) have asymptotically consistent and efficient estimators of the fixed effects without assuming normality. Hirst et al. (1991) rely on the asymptotic properties of maximum likelihood estimators and therefore need to assume normality. This also puts Hirst et al. (1991) in a position to make inferences about the variance components, something that Vonesh and Carter (1992) can't do. However, if the distributional assumptions are not valid, the Hirst et al. (1991) asymptotic theory breaks down. At first glance the nonlinear fixed effects with linear random effects model (2.67) does not appear to be that useful. However, letting $\mathbf{Z}_i$ be a vector of all ones will yield the compound symmetric covariance structure discussed by Jennrich and Schluchter (1986) and others. Letting $\mathbf{Z}_i$ be an identity matrix can yield the unstructured covariance matrix also discussed by Jennrich and Schluchter (1986) and others. Next we will switch back to the model where the random effects are in the nonlinear part of the model.

Lindstrom and Bates (1990) take the approach of expanding model (2.56) about current estimates of the fixed effects ($\boldsymbol{\alpha}_0$) and conditional expectation of random effects ($\mathbf{b}_{i0} = E[\mathbf{b}_i|\mathbf{y}_i]$). The approximate model becomes:

$$\mathbf{y}_i \approx \mathbf{f}(\boldsymbol{\alpha}_0, \mathbf{b}_{i0}, \mathbf{x}_i) + \mathbf{H}_i(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0) + \mathbf{Z}_i(\mathbf{b}_i - \mathbf{b}_{i0}) + \mathbf{e}_i \tag{2.72}$$

where

$$\mathbf{H}_i = \partial \mathbf{f}/\partial \boldsymbol{\alpha}'\Big|_{\substack{\alpha=\alpha_0 \\ b_i=b_{i0}}} \quad \mathbf{Z}_i = \partial \mathbf{f}/\partial \mathbf{b}_i'\Big|_{\substack{\alpha=\alpha_0 \\ b_i=b_{i0}}} . \tag{2.73}$$

This model can also be written as:

$$\mathbf{y}_i - \mathbf{f}(\boldsymbol{\alpha}_0, \mathbf{b}_{i0}, \mathbf{x}_i) + \mathbf{H}_i\boldsymbol{\alpha}_0 + \mathbf{Z}_i\mathbf{b}_{i0} \approx \mathbf{H}_i\boldsymbol{\alpha} + \mathbf{Z}_i\mathbf{b}_i + \mathbf{e}_i \tag{2.74}$$

or

$$\mathbf{y}_i^* \approx \mathbf{H}_i\boldsymbol{\alpha} + \mathbf{Z}_i\mathbf{b}_i + \mathbf{e}_i. \tag{2.75}$$

Lindstrom and Bates (1990) minimize the following objective function:

$$O(\boldsymbol{\alpha}, \mathbf{D}, \sigma^2|\mathbf{y}_i) = \sum_{i=1}^{M} \log\left|\mathbf{Z}_i\mathbf{D}\mathbf{Z}_i' + \sigma^2\mathbf{I}_{n_i}\right| + \sum_{i=1}^{M}[\mathbf{y}_i^* - \mathbf{H}_i\boldsymbol{\alpha}]'(\mathbf{Z}_i\mathbf{D}\mathbf{Z}_i' + \sigma^2\mathbf{I}_{n_i})^{-1}[\mathbf{y}_i^* - \mathbf{H}_i\boldsymbol{\alpha}]. \tag{2.76}$$

Again, with the normality assumption, except for a constant (2.76) is the -2 log likelihood for the approximate model (2.75). This appears to be the same objective function that Hirst et al. (1991) minimize (2.66). The difference here is that the matrices of partial derivatives are being evaluated at current estimates of the fixed effects ($\boldsymbol{\alpha}_0$) and the conditional expectation of the random effects ($\mathbf{b}_{i0}$=E[$\mathbf{b}_i|\mathbf{y}_i^*$]). Also, the $\mathbf{y}_i^*$ is different for this objective function. Lindstrom and Bates (1990) use the above objective function to estimate the variance components. They use a Newton-Raphson (NR) algorithm to accomplish this minimization. They prefer the NR algorithm over the EM algorithm because of speed of convergence and the ability to use the orthogonality convergence criteria (Bates and Watts, 1981). Once estimates of the variance components have been acquired, they use them to obtain updated estimates of the fixed effects ($\boldsymbol{\alpha}_1$) and the conditional expectation of the random effects ($\mathbf{b}_{i1}$=E[$\mathbf{b}_i|\mathbf{y}_i^*$]). They accomplish this by minimizing the following objective function:

$$O(\boldsymbol{\alpha}, \mathbf{b} | \mathbf{y}_i, \sigma^2, \mathbf{D}) = \sum_{i=1}^{M} \tfrac{-1}{2}\sigma^{-2}[\mathbf{y}_i - \mathbf{f}(\boldsymbol{\alpha}, \mathbf{b}_i, \mathbf{x}_i)]'[\mathbf{y}_i - \mathbf{f}(\boldsymbol{\alpha}, \mathbf{b}_i, \mathbf{x}_i)] + \sum_{i=1}^{M} \tfrac{-1}{2}\mathbf{b}_i'\mathbf{D}\mathbf{b}_i \ . \quad (2.77)$$

For a fixed $\boldsymbol{\alpha}$, the above objective function is a constant plus the log of the posterior density of $\mathbf{b}$. Therefore the $\mathbf{b}$ that minimizes (2.77) is the posterior mode. The $\boldsymbol{\alpha}$ that minimizes (2.77) is a maximum likelihood estimator of an approximate marginal distribution of $\mathbf{y}$. These new estimates of the fixed effects and variance components are then used to update the approximate model (2.74) and the whole process is repeated. Lindstrom and Bates (1990) recommend using the Hessian from the last iteration of the approximate model (2.74) to obtain standard errors of fixed effects and variance component estimates. They add the following caveat:

> These uncertainty estimates are approximate in that they are based on a linear approximation to the model function at the parameter estimates. This type of approximation is commonly used to estimate the uncertainty in the parameters of nonlinear models. As pointed out by Bates and Watts (1988, Chap 6), these uncertainty estimates can be quite inaccurate and a better appreciation of the uncertainty can be obtained by evaluating the profile likelihood and creating pairwise plots of the projected likelihood contours.

The big disadvantage of using the Lindstrom and Bates (1990) method is that minimizing the objective function (2.77) can quickly become a cumbersome task. If there are 100 individuals and 3 random effects, then (2.77) will have to be minimized with respect to at least 300 parameters. Like Sheiner and Beal (1980) and Hirst et al. (1991), Lindstrom and Bates (1990) do not require that each subject have a minimum amount of data. Vonesh and Carter (1992) complain that Lindstrom and Bates' (1990) estimates have, "as yet, unspecified asymptotic properties." To the contrary, Lindstrom and Bates (1990)

have maximum likelihood estimates for the approximate model (2.75). Therefore, for the approximate model, the estimates are asymptotically consistent and efficient. Lindstrom and Bates (1990) also show how their method can be used to derive restricted maximum likelihood estimates.

Following the lead of Lindstrom and Bates (1990), Young et al. (1992) obtains the same approximate model (2.75). However, unlike Lindstrom and Bates (1990), he recommends using the EM algorithm to minimize the -2 log likelihood function (2.76). Inherent in the EM algorithm is the ability to estimate $E(\mathbf{b}_i|\mathbf{y}_i)$, once the approximate model (2.75) converges, the new estimates for $\boldsymbol{\alpha}$ and $E(\mathbf{b}_i|\mathbf{y}_i)$ are used to update $\mathbf{H}_i$, $\mathbf{Z}_i$, and $\mathbf{y}_i^*$ and the whole process is repeated until the -2 log likelihood of the approximate model converges to a minimum. This method differs from Lindstrom and Bates (1990) in that the EM algorithm is used to obtain the most current estimates of the fixed effects and conditional expectation of the random effects whereas Lindstrom and Bates use another iterative scheme of minimizing (2.77). In some sense, using (2.77) brings Lindstrom and Bates back closer to the original model (2.56), however they suggest using the approximate model (2.75) for making inferences about the fixed effects and variance components. Pinheiro and Bates (1995) indicate that using the method suggested by Young et al. (1992) will result in convergence to the same estimates as those obtained by employing the Lindstrom and Bates (1990) method. In fact, using the method suggested by Young et al. (1992) the "Orange Tree" example published by Lindstrom and Bates (1990) has been replicated. At convergence, the method suggested

by Young et al. (1992) yields maximum likelihood estimates for the approximate model (2.75). It is also possible to use the EM algorithm to obtain RML estimates.

In a recent article, Pinheiro and Bates (1995) compare the method of Lindstrom and Bates (1990), with 3 other more numerically intensive approximations of the likelihood equation (2.52). They conclude the following:

> We conclude that the linear mixed-effects (LME) approximation suggested by Lindstrom and Bates, the Laplacian approximation, and Guassian quadrature centered at the conditional modes of the random effects are quite accurate and computationally efficient.

It has been shown that all of the methods discussed in this thesis for estimating model (2.56) involve some sort of first order Taylor series expansion. The different methods essentially relate to different ways to accomplish this Taylor series expansion. For all methods, the resulting asymptotic theory estimates are only as good as the original linear approximation to the nonlinear model. When the original model possesses a problem known as curvature, the results obtained from the application of asymptotic theory of the linearized model can be very misleading (Seber and Wild, 1989). Problems associated with nonlinear models, including curvature will be briefly discussed in Chapter V.

As of yet, the nonlinear mixed effects model has not been extended to handle the case of multi-response data. The development of such models will be the subject of the next chapter.

CHAPTER III

SOME MODELS

In this chapter two models for analyzing multi-response nonlinear longitudinal data will be presented. The first model combines the multi-response aspect of Zucker et al. (in press) with the nonlinear aspect discussed by Sheiner and Beal (1980), Lindstrom and Bates (1990), Hirst et al. (1991) and Vonesh and Carter (1992).

3.1 A Multi-response Nonlinear Mixed Effects Model with Nonlinear Random Effects

Consider the following subject-time-response specific model:

$$y_{ijk} = f_k(\underset{p_k\times 1}{\boldsymbol{\alpha}_k}, \underset{q_k\times 1}{\mathbf{b}_{ik}}, x_{ij}) + e_{ijk}. \tag{3.1}$$

The y_{ijk} is the outcome for the kth response variable measured on the ith individual at the jth time (i=1,...,M, j=1,...,m_i, and k=1,...,r). Where M is the total number of subjects, m_i is the number of repeated measurements for the ith individual, and r is the number of responses. The f_k is a nonlinear function associated with the kth response variable. $\boldsymbol{\alpha}_k$ is an unknown vector of fixed effects associated with the kth response that needs to be estimated. $\mathbf{b}_{ik}$ is a vector of random effects for the ith individual's kth response function. The x_{ij} is the subject-time specific covariate. This covariate is usually dose or time. The e_{ijk} is a random error term. If the different subject-time specific responses are stacked one on top of each other the resulting subject-time specific model can be written as:

$$\begin{bmatrix} y_{ij1} \\ y_{ij2} \\ \vdots \\ y_{ijr} \end{bmatrix} = \begin{bmatrix} f_1(\boldsymbol{\alpha}_1, \mathbf{b}_{i1}, x_{ij}) \\ f_2(\boldsymbol{\alpha}_2, \mathbf{b}_{i2}, x_{ij}) \\ \vdots \\ f_r(\boldsymbol{\alpha}_r, \mathbf{b}_{ir}, x_{ij}) \end{bmatrix} + \begin{bmatrix} e_{ij1} \\ e_{ij2} \\ \vdots \\ e_{ijr} \end{bmatrix}. \tag{3.2}$$

In this model, it will be assumed that:

$$\mathbf{b}_i = \begin{bmatrix} \mathbf{b}_{i1} \\ \mathbf{b}_{i2} \\ \vdots \\ \mathbf{b}_{ir} \end{bmatrix} \sim N(\mathbf{0}, \underset{q\times q}{\mathbf{D}}) \text{ where } q = \sum_{k=1}^{r} q_k \tag{3.3}$$

and

$$\mathbf{e}_{ij} = \begin{bmatrix} e_{ij1} \\ e_{ij2} \\ \vdots \\ e_{ijr} \end{bmatrix} \sim N(\mathbf{0}, \underset{r\times r}{\mathbf{S}}) . \tag{3.4}$$

The $\mathbf{b}_i$ and $\mathbf{e}_{ij}$ are also assumed to be independent. For notational ease model (3.2) can be written as:

$$\underset{\sim}{\mathbf{y}}_{ij} = \underline{\mathbf{F}}(\boldsymbol{\alpha}, \mathbf{b}_i, x_{ij}) + \mathbf{e}_{ij} . \tag{3.5}$$

If it were not for the possibility of missing data, model (3.5) would be adequate. At each subject-time specific point there is the possibility that one or more of the multiple responses is missing. Thus, in estimating the within subject error structure it is important

to keep track of where the missing responses occur. The following modification to model (3.5) is proposed:

$$\underset{n_{ij}\times 1}{\mathbf{y}_{ij}} = \mathbf{F}(\boldsymbol{\alpha}, \mathbf{b}_i, x_{ij}) + \underset{n_{ij}\times r}{\mathbf{U}_{ij}}\,\underset{r\times 1}{\mathbf{e}_{ij}}. \tag{3.6}$$

In this formulation n_{ij} is the number of responses being measured at the jth time point for the ith individual. If no responses are missing then $n_{ij}=r$. $\mathbf{U}_{ij}$ is an indicator matrix. A simple example may help to clarify the structure of $\mathbf{U}_{ij}$.

Example: Suppose a researcher is interested in measuring growth in children. At ages 8, 8.5, 9, and 9.5 the researcher measures head circumference, height, and weight. Suppose that at age 9 one child's height was not recorded (reason unknown). Then for that child:

$$\mathbf{U}_{i3} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The $\mathbf{U}_{ij}$ matrix can be used to incorporate non-time dependent covariables, such as intelligence (IQ), into the model. The covariable is added to the response vector. Not wanting to model the within subject error of the covariable, the corresponding row of $\mathbf{U}_{ij}$ is set to all zeros. This makes it possible to obtain correlations between non-time dependent covariables and parameters in the nonlinear model.

Estimation of $\boldsymbol{\alpha}$, $\mathbf{b}_i$, $\mathbf{S}$, and $\mathbf{D}$ via linearization will be discussed in the next section.

3.2 Linearizing the Nonlinear Model

Following the lead of Lindstrom and Bates (1990), model (3.6) will be linearized using Taylor series expansion about $\boldsymbol{\alpha}_0$ and $\mathbf{b}_{i0}$ (initial estimates of $\boldsymbol{\alpha}$ and $E[\mathbf{b}_i|\mathbf{y}_i]$). Doing so results in the following:

$$\mathbf{y}_{ij} - \mathbf{F}(\boldsymbol{\alpha}_0, \mathbf{b}_{i0}, x_{ij}) \approx \mathbf{H}_{ij}(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0) + \mathbf{Z}_{ij}(\mathbf{b}_i - \mathbf{b}_{i0}) + \mathbf{U}_{ij}\mathbf{e}_{ij}. \tag{3.7}$$

$\mathbf{H}_{ij}$ and $\mathbf{Z}_{ij}$ are matrices of partial derivatives of $\mathbf{F}$ with respect to $\boldsymbol{\alpha}$ and $\mathbf{b}_i$ evaluated at the initial estimates and the value of the covariate. The model (3.7) can be written as:

$$\mathbf{y}_{ij} - \mathbf{F}(\boldsymbol{\alpha}_0, \mathbf{b}_{i0}, x_{ij}) + \mathbf{H}_{ij}\boldsymbol{\alpha}_0 + \mathbf{Z}_{ij}\mathbf{b}_{i0} \approx \mathbf{H}_{ij}\boldsymbol{\alpha} + \mathbf{Z}_{ij}\mathbf{b}_i + \mathbf{U}_{ij}\mathbf{e}_{ij} \tag{3.8}$$

or

$$\mathbf{y}^*_{ij} \approx \mathbf{H}_{ij}\boldsymbol{\alpha} + \mathbf{Z}_{ij}\mathbf{b}_i + \mathbf{U}_{ij}\mathbf{e}_{ij}. \tag{3.9}$$

The model can now be written in the following subject specific form:

$$\mathbf{y}^*_i \approx \mathbf{H}_i\boldsymbol{\alpha} + \mathbf{Z}_i\mathbf{b}_i + \mathbf{U}_i\mathbf{e}_i \tag{3.10}$$

where

$$\mathbf{y}^*_i = \begin{bmatrix} \mathbf{y}^*_{i1} \\ \vdots \\ \mathbf{y}^*_{im_i} \end{bmatrix}, \mathbf{H}_i = \begin{bmatrix} \mathbf{H}_{i1} \\ \vdots \\ \mathbf{H}_{im_i} \end{bmatrix}, \mathbf{Z}_i = \begin{bmatrix} \mathbf{Z}_{i1} \\ \vdots \\ \mathbf{Z}_{im_i} \end{bmatrix}, \mathbf{e}_i = \begin{bmatrix} \mathbf{e}_{i1} \\ \vdots \\ \mathbf{e}_{im_i} \end{bmatrix} \text{ and } \mathbf{U}_i = \begin{bmatrix} \mathbf{U}_{i1} & & \\ & \ddots & \\ & & \mathbf{U}_{im_i} \end{bmatrix}.$$

The subject specific model (3.10) is very similar to the model discussed by Laird and Ware (1982). Given initial estimates $\boldsymbol{\alpha}_0$ and $\boldsymbol{\theta}_0$={unique elements of $\mathbf{S}$ and $\mathbf{D}$}, the EM

algorithm can be used to obtain updated estimates of $\boldsymbol{\alpha}$ and $E(\mathbf{b}_i|\mathbf{y}_i^*)$ for the approximate model (3.8). These new estimates are then used to update $\mathbf{H}_i$, $\mathbf{Z}_i$, and $\mathbf{y}_i^*$ in the approximate model (3.8) and then the estimation process is repeated. Model (3.8) is a first-order approximation of model (3.6). Had model (3.6) been expanded around $\mathbf{b}_{i0}=E(\mathbf{b}_i)=\mathbf{0}$ then (3.8) would become:

$$\mathbf{y}_{ij} - \mathbf{F}(\boldsymbol{\alpha}_0, \mathbf{b}_{i0} = \mathbf{0}, x_{ij}) + \mathbf{H}_{ij}\boldsymbol{\alpha}_0 \approx \mathbf{H}_{ij}\boldsymbol{\alpha} + \mathbf{Z}_{ij}\mathbf{b}_i + \mathbf{U}_{ij}\mathbf{e}_{ij}. \tag{3.11}$$

This model (3.11) would be more in line with the methods proposed by Sheiner and Beal (1980), Hirst et al. (1991), and Vonesh and Carter (1992). Using the method suggested by Young et al. (1992) in model (3.8), the $\mathbf{b}_{i0}$ are <u>not</u> set equal to their expectation but rather to their conditional expectation given the data (i.e. $\mathbf{b}_{i0}=E[\mathbf{b}_i|\mathbf{y}_i^*]$). This conditional expectation is a byproduct of the EM algorithm and the methodology is more in line with the methods proposed by Lindstrom and Bates (1990). With this in mind the EM algorithm will be discussed next.

3.3 Use of the EM Algorithm in MNLMEM

The use of the EM algorithm for maximum likelihood estimation has been discussed by Dempster et al. (1977). For each iteration of the EM algorithm, given current estimates of $\mathbf{S}$ and $\mathbf{D}$, $\boldsymbol{\alpha}$ is estimated by:

$$\hat{\boldsymbol{\alpha}} = \left(\sum_{i=1}^{M} \mathbf{H}_i'\mathbf{W}_i\mathbf{H}_i\right)^{-1}\left(\sum_{i=1}^{M} \mathbf{H}_i'\mathbf{W}_i\mathbf{y}_i^*\right). \tag{3.12}$$

The individual random effects can be estimated as:

$$\hat{\mathbf{b}}_i = \mathbf{D}\mathbf{M}_i'\mathbf{W}_i(\mathbf{y}_i^* - \mathbf{H}_i\hat{\boldsymbol{\alpha}}). \tag{3.13}$$

where $\mathbf{W}_i = \mathbf{V}_i^{-1}(\boldsymbol{\theta})$ and $\mathbf{V}_i(\boldsymbol{\theta}) = \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i' + \mathbf{U}_i(\mathbf{I} \otimes \mathbf{S})\mathbf{U}_i'$. To use the EM algorithm to estimate the variance components, we consider the "complete" data. If it were possible to observe the $\mathbf{b}_i$ and $\mathbf{e}_i$ in addition to the $\mathbf{y}_i^*$ then it would be easy to write down closed form maximum likelihood solutions for $\boldsymbol{\theta}$. If $\mathbf{S}$ and $\mathbf{D}$ are arbitrary, the complete data maximum likelihood estimate of $\mathbf{S}$ is given by:

$$\mathbf{S} = \left(\sum_{i=1}^{M}\sum_{j=1}^{m_i}\mathbf{e}_{ij}\mathbf{e}_{ij}'\right)/n = \mathbf{T}_1/n \text{ where } n = \Sigma_{i=1}^{M} m_i. \tag{3.14}$$

The complete data maximum likelihood estimate for $\mathbf{D}$ is given by:

$$\mathbf{D} = \sum_{i=1}^{M}\mathbf{b}_i\mathbf{b}_i'/M = \mathbf{T}_2/M. \tag{3.15}$$

In (3.14) and (3.15) $\mathbf{T}_1$ and $\mathbf{T}_2$ are the sufficient statistics for $\boldsymbol{\theta}$. Since the "complete" data are not available, estimation of the sufficient statistics is necessary. This is carried out by taking the expected value of an estimate of the sufficient statistics conditional on the observed data ($\mathbf{y}^*$). Hence, the name E-step is used for taking the expected value. To estimate the sufficient statistic $\mathbf{T}_1$, the following expectation is taken:

$$\hat{\mathbf{T}}_1 = E\left\{\sum_{i=1}^{M}\sum_{j=1}^{m_i}\mathbf{e}_{ij}\mathbf{e}_{ij}'|\mathbf{y}_i^*, \hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\theta}}), \hat{\boldsymbol{\theta}}\right\} \tag{3.16}$$

$$= \sum_{i=1}^{M}\sum_{j=1}^{m_i}\{\mathbf{V}[\mathbf{e}_{ij}|\mathbf{y}_i^*, \hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\theta}}), \hat{\boldsymbol{\theta}}] + \mathbf{E}[\mathbf{e}_{ij}|\mathbf{y}_i^*, \hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\theta}}), \hat{\boldsymbol{\theta}}]\mathbf{E}[\mathbf{e}_{ij}|\mathbf{y}_i^*, \hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\theta}}), \hat{\boldsymbol{\theta}}]'\} \tag{3.17}$$

$$= \sum_{i=1}^{M}\sum_{j=1}^{m_i}[\mathbf{R}_i]_j. \tag{3.18}$$

Where $[\mathbf{R}_i]_j$ is the jth $r \times r$ block of the block diagonal matrix $\mathbf{R}_i$ which is defined as follows:

$$\underset{rm_i \times rm_i}{\mathbf{R}_i} = \mathbf{V}[\mathbf{e}_i|\mathbf{y}_i^*, \hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\theta}}), \hat{\boldsymbol{\theta}}] + \mathbf{E}[\mathbf{e}_i|\mathbf{y}_i^*, \hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\theta}}), \hat{\boldsymbol{\theta}}]\mathbf{E}[\mathbf{e}_i|\mathbf{y}_i^*, \hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\theta}}), \hat{\boldsymbol{\theta}}]' \tag{3.19}$$

where

$$\mathbf{V}[\mathbf{e}_i|\mathbf{y}_i^*, \hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\theta}}), \hat{\boldsymbol{\theta}}] = (\mathbf{I}_{m_i} \otimes \mathbf{S}) - (\mathbf{I}_{m_i} \otimes \mathbf{S})\mathbf{U}_i'\mathbf{W}_i\mathbf{U}_i(\mathbf{I}_{m_i} \otimes \mathbf{S}) \tag{3.20}$$

and

$$\mathbf{E}[\mathbf{e}_i|\mathbf{y}_i^*, \hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\theta}}), \hat{\boldsymbol{\theta}}] = (\mathbf{I}_{m_i} \otimes \mathbf{S})\mathbf{U}_i'\mathbf{W}_i(\mathbf{y}_i^* - \mathbf{H}_i\boldsymbol{\alpha}). \tag{3.21}$$

To estimate the sufficient statistic $\mathbf{T}_2$ the following expectation is taken:

$$\hat{\mathbf{T}}_2 = E\left\{\sum_{i=1}^{M}\mathbf{b}_i\mathbf{b}_i'|\mathbf{y}_i^*, \hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\theta}}), \hat{\boldsymbol{\theta}}\right\} \tag{3.22}$$

$$= \sum_{i=1}^{M}\{\mathbf{V}[\mathbf{b}_i|\mathbf{y}_i^*, \hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\theta}}), \hat{\boldsymbol{\theta}}] + \mathbf{E}[\mathbf{b}_i|\mathbf{y}_i^*, \hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\theta}}), \hat{\boldsymbol{\theta}}]\mathbf{E}[\mathbf{b}_i|\mathbf{y}_i^*, \hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\theta}}), \hat{\boldsymbol{\theta}}]'\} \tag{3.23}$$

where

$$\mathbf{V}[\mathbf{b}_i|\mathbf{y}_i^*, \hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\theta}}), \hat{\boldsymbol{\theta}}] = \mathbf{D} - \mathbf{D}\mathbf{Z}_i'\mathbf{W}_i\mathbf{Z}_i\mathbf{D} \tag{3.24}$$

and

$$\mathbf{E}[\mathbf{b}_i|\mathbf{y}_i^*, \hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\theta}}), \hat{\boldsymbol{\theta}}] = \mathbf{D}\mathbf{Z}_i'\mathbf{W}_i\{\mathbf{y}_i^* - \mathbf{H}_i\boldsymbol{\alpha}\}. \tag{3.25}$$

In (3.16 through 3.25) **D**, **S**, and $\boldsymbol{\alpha}$ are based on the previous M-step of the EM algorithm. The next step in the EM algorithm is to obtain the maximum likelihood estimates from the sufficient statistics. This is accomplished as:

$$\hat{\mathbf{S}} = \mathbf{T}_1/n \tag{3.26}$$

and

$$\hat{\mathbf{D}} = \mathbf{T}_2/M. \tag{3.27}$$

Now the estimation process (3.12, 3.13, 3.26, and 3.27) can be repeated. During the EM algorithm the -2 log likelihood for the approximate model (3.10) can be calculated as:

$$-2\ln L = N\ln(2\pi) + \sum_{i=1}^{M} \ln(|\mathbf{V}_i(\boldsymbol{\theta})| + \sum_{i=1}^{M} (\mathbf{y}_i^* - \mathbf{H}_i\boldsymbol{\alpha})'\mathbf{V}_i^{-1}(\boldsymbol{\theta})(\mathbf{y}_i^* - \mathbf{H}_i\boldsymbol{\alpha}). \tag{3.28}$$

In (3.28) N is the total number of observations. When the change in the -2 log likelihood is sufficiently small, the process is considered to have converged. At this point, the estimate of $\boldsymbol{\alpha}$ is used to update the initial estimate $\boldsymbol{\alpha}_0$ and the conditional expectation of the random effects is used to update the $\mathbf{b}_{i0}$ in the approximate model (3.8) and the process is then repeated.

Dempster et al. (1977) also discuss restricted maximum likelihood estimation (RML). This can be accomplished by replacing $\mathbf{W}_i$ with the following formulation in (3.20, 3.21, 3.24, and 3.25):

$$\mathbf{W}_i = \mathbf{V}_i^{-1} - \mathbf{V}_i^{-1}\mathbf{H}_i(\mathbf{H}_i'\mathbf{V}_i^{-1}\mathbf{H}_i)^{-1}\mathbf{H}_i'\mathbf{V}_i^{-1}. \tag{3.29}$$

The likelihood can be calculated as:

$$-2\ln L_{REML} = -2\ln L_{ML} + \log\left(\left|\sum_{i=1}^{M} \mathbf{H}_i' \mathbf{V}_i^{-1} \mathbf{H}_i\right|\right). \tag{3.30}$$

As previously discussed in section 2.10 there are some problems when the random effects enter the model nonlinearly. The problem being that the likelihood surface can not be described well and an approximation to the likelihood surface must be used. One way to avoid this problem is to only consider models with additive errors. Such a model will be the focus of the next section.

3.4 A Multi-response Nonlinear Mixed Effects Model with Linear Covariance Structure

Consider the following multi-response nonlinear mixed effects model:

$$y_{ijk} = f_k(\boldsymbol{\alpha}_k, x_{ij}) + e_{ijk}. \tag{3.31}$$

The y_{ijk} is the outcome for the kth response variable measured on the ith individual at the jth time (i=1,...,M, j=1,...,m_{ik}, and k=1,...,r). Where M is the total number of subjects, m_{ik} is the number of repeated measurements for the ith individual's kth response, and r is the number of multiple responses. The f_k is a nonlinear function associated with the kth response variable. The x_{ij} is a fixed covariate associated with the jth measurement on the ith individual. In medical research this covariate is often time or dose. $\boldsymbol{\alpha}_k$ is a $p_k \times 1$

vector of unknown fixed effects associated with the kth response variable. The e_{ijk} is the random error. The complete vector of random errors will be assumed to have the following distribution:

$$\mathbf{e} \sim N[\mathbf{0}, \mathbf{V}(\boldsymbol{\theta})] \quad \text{where} \quad \mathbf{V}(\boldsymbol{\theta}) = \sum_q \theta_q \mathbf{V}_q. \tag{3.32}$$

The random effects are entered into the model by specifying the $\mathbf{V}_q$ and estimating θ_q. As discussed next, this model is <u>not</u> a special case of the MNLMEM discussed in sections 3.1 through 3.3.

3.5 Comparison of both MNLMEMs

At first glance, model (3.31) looks like a special case of model (3.1) where there are no random effects. However, it will be shown that the error structure for model (3.31) is more general than the special case of model (3.1). Working with the subject-time specific version (3.6) of model (3.1), and assuming that there are no random effects, it can be shown that the model is:

$$\underset{n_{ij}\times 1}{\mathbf{y}_{ij}} \sim N[\underset{n_{ij}\times 1}{\mathbf{F}(\boldsymbol{\alpha}, x_{ij})}, \underset{n_{ij}\times r}{\mathbf{U}_{ij}} \underset{r\times r}{\mathbf{S}} \underset{r\times n_{ij}}{\mathbf{U}_{ij}'}]. \tag{3.33}$$

If model (3.31) were to be written in a subject-time specific format it could be shown that:

$$\underset{n_{ij}\times 1}{\mathbf{y}_{ij}} \sim N[\underset{n_{ij}\times 1}{\mathbf{F}(\boldsymbol{\alpha}, x_{ij})}, \underset{n_{ij}\times n_{ij}}{\mathbf{V}(\boldsymbol{\theta}) = \sum_q \theta_q \mathbf{V}_q}]. \tag{3.34}$$

The only difference in the two models is the variance structure. As used in sections 3.1 through 3.3, the matrix $\mathbf{U}_{ij}$ in (3.33) is just an indicator matrix. Thus if there are no data missing, $\mathbf{U}_{ij}=\mathbf{I}_r$. In this sense, the covariance structure listed in (3.33) is the multivariate analog of $\sigma^2\mathbf{I}$. Even if the $\mathbf{U}_{ij}$ are used to structure the covariance the following has been shown by Zerbe et al. (1994):

$$\mathbf{U}_{ij}\mathbf{S}\mathbf{U}_{ij}' = \sum_{t=1}^{r}\sum_{u=t}^{r}[S]_{tu}\left[[\mathbf{U}_{ij}]_t[\mathbf{U}_{ij}]_u' + (1-\delta_{tu})[\mathbf{U}_{ij}]_t[\mathbf{U}_{ij}]_u'\right] = \sum_q \theta_q \mathbf{V}_q . \qquad (3.35)$$

The $[S]_{tu}$ is the (t,u)th element of the matrix $\mathbf{S}$, the $[\mathbf{U}_{ij}]_u$ is the uth column of $\mathbf{U}_{ij}$, and δ_{tu} is unity when $t=u$ and zero otherwise. In other words, the special case of model (3.1) where there are no random effects in the model, is in fact a special case of model (3.31). In a similar fashion, the special case of model (3.1) where random effects enter the model linearly, can also be shown to be special case of model (3.31). The "banded" (Jennrich and Schluchter, 1986) error structure, is an example of a covariance structure that can be picked up by model (3.31) but <u>not</u> by model (3.1). It should be noted that the use of $\mathbf{U}_{ij}$ in sections 3.1 through 3.3 makes the handling of missing data easy. This is not the case with model (3.31) where the handling of missing data requires that close attention be paid to the construction of variance design matrices $\mathbf{V}_q$. Next estimation of the parameters in the linear covariance structure model will be discussed.

3.6. Estimation in the MNLMEM with Linear Covariance Structure

Now the task becomes to estimate the fixed effects ($\boldsymbol{\alpha}_k$) and the variance components (θ_q). The nonlinear part of model (3.31) can be "linearized" by using Taylor series expansion about an initial guess $\boldsymbol{\alpha}_k^{(0)}$. The model becomes:

$$(y_{ijk} - f_k(x_{ijk}, \boldsymbol{\alpha}_k^{(0)})) \approx \mathbf{H}_{ijk}(\boldsymbol{\alpha}_k - \boldsymbol{\alpha}_k^{(0)}) + e_{ijk}. \tag{3.36}$$

Where $\mathbf{H}_{ijk} = \left[\frac{\partial f_k}{\partial \boldsymbol{\alpha}_k}\right]_{\boldsymbol{\alpha}_k = \boldsymbol{\alpha}_k^{(0)}}$ is a matrix of partial derivatives evaluated at $\boldsymbol{\alpha}_k^{(0)}$ and the covariable. When putting the data into the format of model (3.37) it is often easiest to stack <u>all</u> the individual observations involving one response variable on top of <u>all</u> the other individual observations involving the other response variables. Doing so makes it easier to specify the $\mathbf{V}_q$ matrices especially when the data are balanced. Stacking the individual observations on top of each other as follows,

$$\mathbf{y} = [y_{111}, \ldots y_{1m_{11}1} | \ldots | y_{M11}, \ldots y_{Mm_{M1}1} | y_{112}, \ldots y_{1m_{12}2} | \ldots | y_{M12}, \ldots y_{Mm_{M2}2} | \ldots | y_{11r}, \ldots y_{1m_{1r}r} | \ldots | y_{M1r}, \ldots y_{Mm_{Mr}r}]'$$

and linearizing, the complete model becomes:

$$\mathbf{y}^* \approx \mathbf{H}\boldsymbol{\alpha}^* + \mathbf{e}, \quad \text{where} \quad \mathbf{e} \sim N[\mathbf{0}, \mathbf{V}(\boldsymbol{\theta})]. \tag{3.37}$$

Model (3.37) is a special case of a model discussed by Magnus and Neudecker (1988) with a covariance structure of the type discussed by Hocking (1985) and others. Once the model is in this format, it is simply a matter of applying Hocking's algorithm (1985, p.239) to solve for $\boldsymbol{\alpha}^*$ (hence $\boldsymbol{\alpha}$) and θ_q. Hocking's algorithm takes advantage of the

linear structure of the variance and minimizes model (3.37)'s -2 log likelihood (same as maximizing the likelihood):

$$-2\ln L(\boldsymbol{\alpha}^*, \boldsymbol{\theta}) = N\ln(2\pi) + \ln(|\mathbf{V}(\boldsymbol{\theta})|) + \{(\mathbf{y}^* - \mathbf{H}\boldsymbol{\alpha}^*)'\mathbf{V}^{-1}(\boldsymbol{\theta})(\mathbf{y}^* - \mathbf{H}\boldsymbol{\alpha}^*)\}. \quad (3.38)$$

The "new" estimates obtained from minimizing (3.38) can then be used as updated initial values in model (3.36) and the estimation process repeated until the estimates stabilize. Once estimates have been obtained, it is usually desirable to use these estimates to make inferences. In the next section some asymptotic properties for $\boldsymbol{\alpha}$ and $\boldsymbol{\theta}$ will be discussed.

3.7 Asymptotic Properties

Approximate covariances for the estimated parameters $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\theta}}$ may be obtained after convergence of model (3.37) by examining Fisher's information matrix. These covariances are approximate because they are based on the linearization of model (3.31) and not the nonlinear model itself. For the linear model Hocking (1985) shows that under suitable regularity conditions, asymptotically the following holds:

$$\hat{\boldsymbol{\alpha}} \sim N[\boldsymbol{\alpha}, \hat{\mathbf{V}}(\boldsymbol{\alpha}) \approx (\mathbf{H}'\mathbf{V}^{-1}(\hat{\boldsymbol{\theta}})\mathbf{H})^{-1}] \quad (3.39)$$

and independently

$$\hat{\boldsymbol{\theta}} \sim N[\boldsymbol{\theta}, \hat{\mathbf{V}}(\boldsymbol{\theta}) \approx 2\hat{\boldsymbol{\Omega}}^{-1}]. \quad (3.40)$$

The (i,j)th element of $\hat{\boldsymbol{\Omega}}$ is defined as:

$$\hat{\Omega}_{ij} = trace[\mathbf{V}^{-1}(\hat{\boldsymbol{\theta}})\mathbf{V}_i\mathbf{V}^{-1}(\hat{\boldsymbol{\theta}})\mathbf{V}_j]. \quad (3.41)$$

Thus, based on asymptotic theory it is possible to use a Wald test to perform the following hypothesis test:

$$H_0 : \underset{d\times pp\times 1}{\mathbf{C}}\ \boldsymbol{\alpha} = \mathbf{0}.$$

Here $\mathbf{C}$ is a known matrix of full rank. The test statistic is:

$$(\mathbf{C}\hat{\boldsymbol{\alpha}})'[\mathbf{C}\hat{\mathbf{V}}(\boldsymbol{\alpha})\mathbf{C}']^{-1}(\mathbf{C}\hat{\boldsymbol{\alpha}}) \sim \chi^2_d \ . \tag{3.42}$$

In a similar fashion hypothesis tests for $\boldsymbol{\theta}$ can also be conducted. Another possibility for hypothesis testing is to use a likelihood ratio test. It is well known that the change in -2 log likelihood for two "nested" models is chi-square with degrees of freedom equal to the difference in the number of parameters being estimated in the two models. By "nested" it is meant that one model is a special case of the other model. For example, suppose the data have two distinct groups such as male and female. One approach would be to fit a model to each group. Each group would have the same functional form, but the parameters would be allowed to differ. The nested model would ignore the groups and fit a model to the entire data. The functional form of the "no group" model should be the same as used in the two group approach. The "no group" model is "nested" in the two group model because it can be viewed as a special case of the two group model where the parameters are exactly the same for both groups. The group effect can now be tested by computing the change in -2 log likelihood between the two group model and

the no group model. When possible, this likelihood ratio based method of testing is preferable to the Wald test because it has more power.

Frequently in nonlinear longitudinal data interest may not only lie in the parameters themselves, but in nonlinear functions of the parameters. In the next section asymptotic theory for nonlinear functions of the parameters will be discussed.

3.8 Functions of the Parameters

It is often the case for longitudinal models that interest lies in some nonlinear function, say $g(\boldsymbol{\alpha}, \boldsymbol{\theta})$, of the parameters in the model (3.31). Following the work of Zerbe et al. (1994), the nonlinear function g can be expanded around the estimates $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\theta}}$ in a Taylor series fashion to obtain:

$$g(\boldsymbol{\alpha}, \boldsymbol{\theta}) \approx g(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\theta}}) + [\partial g/\partial \boldsymbol{\alpha}|_{\alpha=\hat{\alpha}}]'(\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}) + [\partial g/\partial \boldsymbol{\theta}|_{\theta=\hat{\theta}}]'(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}). \qquad (3.43)$$

This can be rewritten as:

$$g(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\theta}}) \approx g(\boldsymbol{\alpha}, \boldsymbol{\theta}) + [\partial g/\partial \boldsymbol{\alpha}|_{\alpha=\hat{\alpha}}]'(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) + [\partial g/\partial \boldsymbol{\theta}|_{\theta=\hat{\theta}}]'(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}). \qquad (3.44)$$

Thus, based on the Taylor series approximation and the asymptotic properties (3.39) and (3.40) the nonlinear function has an approximately normal distribution with:

$$E[g(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\theta}})] \approx g(\boldsymbol{\alpha}, \boldsymbol{\theta}) \qquad (3.45)$$

and

$$\mathbf{V}[g(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\theta}})] \approx [\partial g/\partial \boldsymbol{\alpha}|_{\alpha=\hat{\alpha}}]' \hat{\mathbf{V}}(\boldsymbol{\alpha})[\partial g/\partial \boldsymbol{\alpha}|_{\alpha=\hat{\alpha}}] + [\partial g/\partial \boldsymbol{\theta}|_{\theta=\hat{\theta}}]' \hat{\mathbf{V}}(\boldsymbol{\theta})[\partial g/\partial \boldsymbol{\theta}|_{\theta=\hat{\theta}}]. \quad (3.46)$$

Now (3.45) and (3.46) can be used to build approximate confidence intervals for the MNLMEM with additive errors.

Zerbe et al. (1994) show that the usual Laird and Ware (1982) model is a special case of the Hocking (1985) model. Following the same argument used in establishing (3.35), it can be shown that the first order approximate models (3.10 and 3.11) are special cases of the Hocking (1985) model. Thus all of the asymptotic theory developed in sections 3.7 and 3.8 can be applied to the approximate models (3.10 and 3.11). The next chapter will demonstrate some of the methods presented in this chapter.

CHAPTER IV

APPLICATION

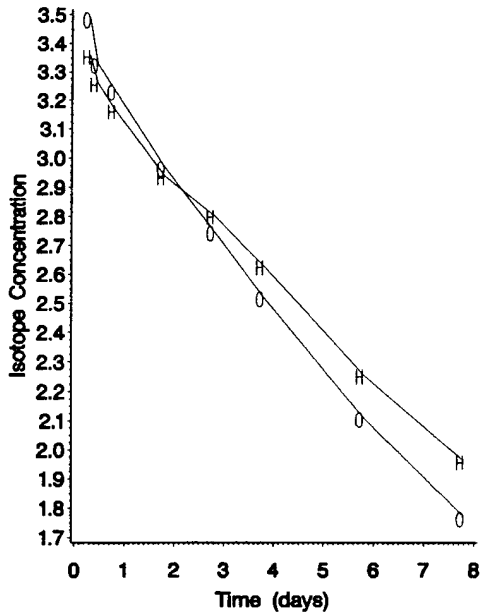
In this chapter three applications of the multi-response nonlinear mixed effects model (MNLMEM) will be discussed. It is not the intent of this section to perform a complete data analysis on the problems presented. The problems are introduced as a means of displaying the virtues of MNLMEM.

4.1 Human Energy Expenditure

One method of measuring energy expenditure in humans has been dubbed the "doubly labeled water technique" (Ravussin et al., 1991). This technique involves administering a dose of doubly labeled water (^2H_2O and $H_2{}^{18}O$) to the patient and then collecting urine samples over a period of time. The outcome is measured as the fraction

Figure 4.1

Plot of a single individual's ^2H and ^{18}O isotope concentrations over time.



of initial dose multiplied by 10^4. This outcome will be referred to as isotope concentration. In patients, the hydrogen isotope will be eliminated as H_2O and the oxygen isotope will be eliminated as H_2O and CO_2. Thus energy expenditure as measured by CO_2 production can essentially be calculated as the difference in the elimination of ^2H and ^{18}O (Schoeller, 1988). As an example the data for a single individual are plotted in Figure 4.1. Data of the type shown in Figure 4.1 can be modeled with the following bivariate nonlinear model:

$$E\begin{pmatrix} y^O \\ y^H \end{pmatrix} = \begin{pmatrix} \alpha^O \cdot \exp(\beta^O \cdot t) \\ \alpha^H \cdot \exp(\beta^H \cdot t) \end{pmatrix}. \tag{4.1}$$

In the above equations the superscript "H" refers to hydrogen and the superscript "O" refers to oxygen. The variable "t" is used to denote time in days. Using methodology discussed by Prentice (1990), Ravussin (1991), and McClatchey (1993), it is possible to show that a measure for energy expenditure can be calculated as a function of the parameters from the above model. An estimate of energy expenditure is:

$$\underset{liters/day}{CO_2} = 101,373\{\beta^H/\alpha^H - \beta^O/\alpha^O\}. \tag{4.2}$$

Ravussin et al. (1991), studied the energy expenditure for 12 individuals. A summary of some of the data that he collected is listed in Table 4.1. One of the purposes of collecting these data is to compare the doubly labeled water technique to a "gold standard" obtained from a respiratory chamber. Traditionally, individual estimates of energy expenditure would be obtained by fitting linear regression lines to log

transformed data. Energy expenditure would then be calculated as a function of the parameters from the linear regression. In addition, frequently only 2 or 3 data points were used to calculate the regression lines (McClatchey, 1993). This approach has several drawbacks. First, the process can be very tedious. For the data set listed in Table 4.1 a total of 24 regression lines would need to be calculated. Secondly, the bivariate aspect of the data is being ignored.

Table 4.1.

Doubly labeled water isotope concentrations over time for 12 individuals.

		Time in Days from Initial Dose									
Subj		0.3	0.45	0.8	1.8	2.8	3.8	4.8	5.8	6.8	7.8
1	O	3.48854	3.32859	3.23474	2.97045	2.74969	2.52143	.	2.11144	.	1.76980
	H	3.35849	3.26164	3.16630	2.93978	2.80495	2.62949	.	2.25653	.	1.96274
2	O	4.84651	4.57692	4.43260	4.07958	3.63036	3.25812	.	2.60030	.	2.10185
	H	4.64355	4.39669	4.30189	3.96784	3.65238	3.35920	.	2.76285	.	2.39742
3	O	3.20667	3.12248	3.04104	2.67324	2.37099	2.10948	1.90107	1.69543	1.50291	1.36215
	H	3.11547	3.04462	2.99848	2.59952	2.41369	2.19710	.	1.84347	.	1.53534
4	O	.	3.81704	3.67215	3.20852	2.72821	2.36556	2.03101	1.76670	.	.
	H	.	3.63285	3.54392	3.15908	2.72974	2.43855	.	1.91005	.	.
5	O	3.98354	3.89723	3.80345	3.40261	3.08061	2.84423	2.61171	2.41170	2.17684	1.99011
	H	3.83665	3.76665	3.61770	3.38561	3.08211	2.87339	2.70545	2.53691	2.33299	2.19331
6	O	3.34111	3.22376	3.16021	2.93201	2.73100	2.54906	.	2.26930	2.10675	1.97775
	H	3.23138	3.14164	3.08666	2.90894	2.71902	2.60510	2.59870	2.44538	2.22306	2.15739
7	O	2.48738	2.41149	2.29184	2.10493	1.90630	1.71684	1.57067	1.38783	1.26306	1.12656
	H	2.34182	2.30957	2.19497	2.05838	1.89827	1.74078	1.62261	1.46549	1.36204	1.25589
8	O	3.63611	.	3.32258	2.89675	3.50003	2.16647	.	1.53860	.	1.11277
	H	3.47015	.	3.22538	2.88807	2.54646	2.26991	1.94030	1.75135	.	1.27193
9	O	2.64678	2.51201	2.37777	2.09786	1.84498	1.61762	.	1.14205	.	.82468
	H	2.56522	2.46954	2.31604	2.09272	1.87358	1.68425	.	1.26456	.	.94117
10	O	4.15769	.	3.74063	3.17787	2.63972	2.15830	.	1.55962	.	1.08578
	H	4.07425	.	3.66860	3.17309	2.70451	2.32274	.	1.69767	.	1.25006
11	O	4.70576	4.58412	4.26297	3.84957	3.38151	3.02387	.	2.34690	.	1.87995
	H	4.52991	4.35606	4.12025	3.83240	3.40579	3.17685	.	2.49886	.	2.08678
12	O	3.80480	3.59962	3.52828	3.15069	2.87971	2.58227	.	2.09486	.	1.71963
	H	3.62541	3.48708	3.32105	3.10180	2.97700	2.69846	.	2.23495	.	1.90039

The data listed in Table 4.1 are perfectly suited for MNLMEM analysis using a stochastic parameter model. Using MNLMEM will provide a unified approach that allows for the

estimation of parameters as well as a platform for hypothesis testing and confidence interval building. Upon examining the plots of each individual it was obvious that a random intercept (α) term was needed and a random decay rate (β) was suggested as well. The full stochastic parameter MNLMEM model fit to the data is:

$$\begin{pmatrix} y_i^O \\ y_i^H \end{pmatrix} = \begin{pmatrix} (\alpha^O + a_i^O)\exp\{(\beta^O + b_i^O)t\} \\ (\alpha^H + a_i^H)\exp\{(\beta^H + b_i^H)t\} \end{pmatrix} + \begin{pmatrix} e_i^O \\ e_i^H \end{pmatrix} . \tag{4.3}$$

In addition, the following distributional assumptions are made:

$$\begin{pmatrix} e_i^O \\ e_i^H \end{pmatrix} \sim N(\mathbf{0}, \underset{2\times2}{\mathbf{S}}) \text{ and } \begin{pmatrix} a_i^O \\ b_i^O \\ a_i^H \\ b_i^H \end{pmatrix} \sim N(\mathbf{0}, \underset{4\times4}{\mathbf{D}}) . \tag{4.4}$$

We can see that in the above model all the parameters vary stochastically across individuals. MNLMEM was used to fit model (4.3), the results are listed in Table 4.2. In addition, MNLMEM was used to fit a model where only the intercepts (a_i^H and a_i^O) where allowed to very across individuals. These results are also listed in Table 4.2. The SAS/IML code used to run the MNLMEM model is listed in appendix A. The iteration history from the maximum likelihood Lindstrom and Bates run is included in appendix B. There is a switch in the program that allows the user to specify whether to use restricted maximum likelihood (RML) or maximum likelihood (ML). The user can also specify whether the Taylor series expansion is done about $E(\mathbf{b}_i|\mathbf{y}_i)$ (i.e. Lindstrom and

Bates [LB] type of expansion) or about $E(\mathbf{b}_i)=\mathbf{0}$ (i.e. Sheiner and Beal (SB) type of expansion).

Table 4.2

Results of fitting stochastic parameter models to the data listed in Table (4.1) using combinations of maximum likelihood (ML), Sheiner and Beal (SB), and Lindstrom and Bates (LB) methodologies.

Method	$\hat{\alpha}^O$	$\hat{\beta}^O$	$\hat{\alpha}^H$	$\hat{\beta}^H$	-2lnL
ML-LB 2 stochastic parameters	3.7217	-0.1112	3.5600	-0.0917	-333.28
ML-SB 4 stochastic parameters	3.7655	-0.1170	3.5964	-0.0965	-616.65
ML-LB 4 stochastic parameters	3.7741	-0.1188	3.6035	-0.0980	-625.27

As suspected, the model with 4 stochastic parameters fits much better than the model with 2 stochastic parameters. Restricted maximum likelihood (RML) estimates were also obtained. However, the RML results differ very little from those reported in Table 4.2 so they are not presented here. Using (4.2), a measure of energy expenditure can be calculated for each individual as:

$$CO_{2i} = 10.1373\hat{\lambda}_i \tag{4.5}$$

where

$$\hat{\lambda}_i = 10^4\{(\hat{\beta}^H + b_i^H)/(\hat{\alpha}^H + a_i^H) - (\hat{\beta}^O + b_i^O)/(\hat{\alpha}^O + a_i^O)\}. \tag{4.6}$$

A nice feature of the E-M algorithm is that the individual parameter estimates are readily available for use in calculating (4.6). Individual mean daily CO_2 production estimates from the MNLMEM model (4.3) are shown in Table 4.3 along with the gold standard

from the respiratory chamber. In addition, Table 4.3 shows the daily CO_2 production for the usual regression approach accomplished by Ravussin (1991).

Table 4.3

MNLMEM estimates of mean daily CO_2 production (in liters per day) using combinations of maximum likelihood (ML), Sheiner and Beal (SB), and Lindstrom and Bates (LB) methodologies. Also, linear regression estimates and the gold standard.

Subject	Gold Standard	Linear Regression	ML-SB 4 stoch. par.	ML-LB 4 stoch. par.
1	499	448	434.30	436.92
2	356	362	328.50	327.41
3	535	517	488.39	495.93
4	393	406	405.97	410.01
5	370	362	381.74	383.29
6	424	417	435.90	436.80
7	711	626	607.[illegible]	616.39
8	480	495	447.75	456.54
9	672	640	619.57	648.13
10	373	384	418.41	425.20
11	332	297	342.81	341.48
12	403	417	411.03	413.48

Table 4.4 compares the various methods with the gold standard using averages and Pearson's correlation coefficient. The methods are comparable; however, only the MNLMEM provides a unified approach that facilitates hypothesis testing and confidence interval building.

Table 4.4

Average CO_2 production and Pearson's correlation coefficients between MNLMEM estimates and gold standard using combinations of maximum likelihood (ML), Sheiner and Beal (SB), and Lindstrom and Bates (LB) methodologies.

Method of Comparison	ML-SB	ML-LB	Linear Regression	Gold Standard
Mean +/- S.D.	443+/-91	449 +/- 91	448 +/- 105	462 +/- 123
Pearson Correlation	0.97	0.96	0.98	1

4.2 Estimating Correlation Coefficients

Using MNLMEM it is possible to estimate nonlinear functions of both the fixed effects and variance components. Once again consider the data plotted in Figure 4.1 and listed in Table 4.1. Suppose the researcher is interested in estimating the following:

$$corr(a_i^O, b_i^H) = cov(a_i^O, b_i^H) / \left[\sqrt{V(a_i^O) V(b_i^H)} \right]. \tag{4.7}$$

Calculating (4.7) is attempting to ascertain whether or not there is a linear relationship between the initial value for oxygen and the decay rate of hydrogen across subjects. This correlation coefficient is simply a nonlinear function of the parameters in the **D** matrix of model (4.3). The estimated **D** matrix from the maximum likelihood Lindstrom and Bates run is:

$$\hat{\mathbf{D}} = \begin{pmatrix} \hat{\theta}_1 & \hat{\theta}_2 & \hat{\theta}_3 & \hat{\theta}_4 \\ \hat{\theta}_2 & \hat{\theta}_5 & \hat{\theta}_6 & \hat{\theta}_7 \\ \hat{\theta}_3 & \hat{\theta}_6 & \hat{\theta}_8 & \hat{\theta}_9 \\ \hat{\theta}_4 & \hat{\theta}_7 & \hat{\theta}_9 & \hat{\theta}_{10} \end{pmatrix} = \begin{pmatrix} 0.50056 & -.00435 & 0.47356 & -.00395 \\ -.00435 & 0.00094 & -.00416 & 0.00087 \\ 0.47356 & -.00416 & 0.4486 & -.00379 \\ -.00395 & 0.00087 & -.00379 & 0.00080 \end{pmatrix}. \tag{4.8}$$

The estimated within subjects error matrix is:

$$\hat{\mathbf{S}} = \begin{pmatrix} \hat{\theta}_{11} & \hat{\theta}_{12} \\ \hat{\theta}_{12} & \hat{\theta}_{13} \end{pmatrix} = \begin{pmatrix} .0015215 & .0013623 \\ .0013623 & .0020208 \end{pmatrix}. \tag{4.9}$$

The estimated correlation coefficient is:

$$\hat{\rho} = \hat{\theta}_4 / \sqrt{\hat{\theta}_1 \hat{\theta}_{10}} = -.00395 / \sqrt{(.50056)(.00080)} = -.1967. \tag{4.10}$$

The above calculations were carried out on a computer using double precision. The results presented may disagree with the results obtained by using a calculator on the rounded numbers.

Using the methods described in chapter 3 it is possible to obtain an estimate of the standard error for (4.10). Putting all of the variance components into a vector $\boldsymbol{\theta}$, it is possible (see chapter 3) to estimate the variance covariance matrix of $\boldsymbol{\theta}$, we will call this matrix $\hat{\mathbf{V}}(\boldsymbol{\theta})$. The correlation coefficient simply becomes a nonlinear function of the elements in $\boldsymbol{\theta}$. That is:

$$\hat{\rho} = g(\hat{\boldsymbol{\theta}}) = \hat{\theta}_4 / \sqrt{\hat{\theta}_1 \hat{\theta}_{10}} \,. \tag{4.11}$$

Following the work of Zerbe et al. (1994), the nonlinear function g can be expanded in a Taylor series about $\boldsymbol{\theta}$. Thus we have:

$$g(\hat{\boldsymbol{\theta}}) \approx g(\boldsymbol{\theta}) + [\partial g / \partial \boldsymbol{\theta}]' \big|_{\theta = \hat{\theta}} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}). \tag{4.12}$$

Then $g(\hat{\boldsymbol{\theta}})$ is asymptotically normally distributed with:

$$E[g(\hat{\boldsymbol{\theta}})] \approx g(\boldsymbol{\theta}) \text{ and } \mathbf{V}[g(\hat{\boldsymbol{\theta}})] \approx [\partial g / \partial \boldsymbol{\theta}]' \hat{\mathbf{V}}(\boldsymbol{\theta}) [\partial g / \partial \boldsymbol{\theta}]. \tag{4.13}$$

Returning to our example we have:

$$dg/d\theta_1 = -\theta_4 / 2\{\sqrt{\theta_1^3 \theta_{10}}\} \tag{4.14}$$

$$dg/d\theta_4 = 1/\sqrt{\theta_1 \theta_{10}} \tag{4.15}$$

$$dg/d\theta_{10} = -\theta_4 / 2\{\sqrt{\theta_{10}^3 \theta_1}\}. \tag{4.16}$$

$\hat{\boldsymbol{\theta}}$ is a 13×1 vector. However, the nonlinear function $g(\hat{\boldsymbol{\theta}})$ involves only 3 parameters. As a result $\partial g/\partial \boldsymbol{\theta}$ will only have 3 nonzero elements. Thus, we can ignore the rows and columns in $\hat{\mathbf{V}}(\boldsymbol{\theta})$ that correspond to the zeros in $\partial g/\partial \boldsymbol{\theta}$. Working with maximum likelihood (ML) and the Lindstrom and Bates methodology the estimate of $\hat{\mathbf{V}}(\boldsymbol{\theta})$ can be found in appendix B. This estimate is obtained from inverting Fisher's information matrix. Thus we have:

$$\mathbf{V}[g(\hat{\boldsymbol{\theta}})] \approx [\partial g/\partial \boldsymbol{\theta}]' \hat{\mathbf{V}}(\boldsymbol{\theta})[\partial g/\partial \boldsymbol{\theta}]$$

$$\approx \begin{pmatrix} 0.1965 & 49.82 & 122.23 \end{pmatrix} \begin{bmatrix} 0.04188 & -.00033 & 2.66E-6 \\ -.00033 & 0.00004 & -5.4E-7 \\ 2.66E-6 & -5.4E-7 & 1.1E-7 \end{bmatrix} \begin{pmatrix} 0.1965 \\ 49.82 \\ 122.23 \end{pmatrix}$$

$$\approx 0.0778 \, . \tag{4.17}$$

Thus an approximate 95% confidence interval based on asymptotic theory is:

$$-.1967 \pm .55 \, . \tag{4.18}$$

Since the confidence interval overlaps 0 there is not enough evidence to support the hypothesis that the higher a patient's initial value for oxygen the lower the patient's decay rate for hydrogen.

4.3 Glucose Tolerance Test

Consider the administration of a glucose tolerance test in humans. Along with the glucose levels being measured, insulin and phosphate levels were also simultaneously measured. Figure 1.1, on page 2, shows the individual profile plots for insulin and phosphate for 13 individuals who received a glucose tolerance test. Each individual was measured exactly 8 times at 30 to 60 minute intervals. The data of Figure 1.1 are listed in Table 4.5.

Table 4.5

Insulin and phosphate levels after administration of a glucose tolerance test for 13 normal individuals.

Subject	INSULIN uU/cl									Subject	PHOSPHATE mg%							
	0	30	60	90	120	180	240	300	minutes		0	30	60	90	120	180	240	300
1	2.4	9.7	10	8.7	7.1	2.9	2.0	2.0		1	4.3	3.3	3.0	2.6	2.2	2.5	3.4	4.4
2	3.0	5.5	9.0	7.0	4.2	3.6	2.0	2.8		2	3.7	2.6	2.6	1.9	2.9	3.2	3.1	3.9
3	1.0	6.0	7.2	5.4	3.6	3.0	2.2	2.5		3	4.0	4.1	3.1	2.3	2.9	3.1	3.9	4.0
4	1.1	4.5	2.8	5.0	3.2	1.1	8.0	1.5		4	3.6	3.0	2.2	2.8	2.9	3.9	3.8	4.0
5	2.0	6.1	9.5	7.0	5.0	3.2	2.5	1.0		5	4.1	3.8	2.1	3.0	3.6	3.4	3.6	3.7
6	1.3	7.2	6.8	4.8	4.0	2.8	1.0	1.5		6	3.8	2.2	2.0	2.6	3.8	3.6	3.0	3.5
7	1.5	8.0	8.4	6.4	3.8	2.2	1.6	1.6		7	3.8	3.0	2.4	2.5	3.1	3.4	3.5	3.7
8	1.0	6.7	7.5	6.0	4.4	1.8	1.2	1.4		8	4.4	3.9	2.8	2.1	3.6	3.8	4.0	3.9
9	3.5	8.4	9.0	10.2	8.7	4.7	3.2	1.5		9	5.0	4.0	3.4	3.4	3.3	3.6	4.0	4.3
10	1.1	4.5	2.8	5.0	3.2	1.1	2.5	1.6		10	3.7	3.1	2.9	2.2	1.5	2.3	2.7	2.8
11	2.4	9.7	10	8.7	7.1	2.9	2.0	2.0		11	3.7	2.6	2.6	2.3	2.9	2.2	3.1	3.9
12	1.0	3.3	3.9	2.5	1.8	1.0	1.0	1.0		12	4.4	3.7	3.1	3.2	3.7	4.3	3.9	4.8
13	4.0	9.1	7.6	3.1	3.0	1.0	1.0	1.0		13	4.7	3.1	3.2	3.3	3.2	4.2	3.7	4.3

Note: Data complements of Dr. Ron Gotlin, Professor of Pediatrics, UCHSC.

Looking at the data plotted in Figure 1.1, there appears to be an inverse relationship between insulin and phosphate. When one goes up the other goes down and viceversa. A researcher may be interested in answering the question; "What is the time lag between

the maximum insulin level and the minimum phosphate level?" The data listed in Table 4.5 are perfectly suited for a multi-response nonlinear mixed effects model (MNLMEM). The following nonlinear functions can be fit to the data listed in Table 4.5 and plotted in Figure 1.1:

$$y^I = f^I(\boldsymbol{\alpha}, t) = \alpha_1 + t^{\alpha_2} \exp\{-t/\alpha_3\} \tag{4.19}$$

and

$$y^P = f^P(\boldsymbol{\alpha}, t) = \alpha_4 + \alpha_5(t+\alpha_6) + \alpha_7/(t+\alpha_6). \tag{4.20}$$

In equations (4.19 and 4.20) the "I" is used to denote insulin, the "P" is used to denote phosphate, and "t" is used to denote time. These models were selected based on plots of average values over time. Also, 0.001 was used for time t=0.

A nice feature of this data set is that there are no missing values. The data are perfectly suited for the MNLMEM with additive errors discussed in section 3.4. The Hocking form (equation 3.37) of the model is:

$$\begin{bmatrix} {}^*\mathbf{y}^I \\ {}^*\mathbf{y}^P \end{bmatrix}_{208\times 1} = \begin{bmatrix} \frac{\partial f^I}{\partial \alpha_1} & \frac{\partial f^I}{\partial \alpha_2} & \frac{\partial f^I}{\partial \alpha_3} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{\partial f^P}{\partial \alpha_4} & \frac{\partial f^P}{\partial \alpha_5} & \frac{\partial f^P}{\partial \alpha_6} & \frac{\partial f^P}{\partial \alpha_7} \end{bmatrix}_{208\times 7} \begin{bmatrix} \alpha_1^* \\ \alpha_2^* \\ \alpha_3^* \\ \alpha_4^* \\ \alpha_5^* \\ \alpha_6^* \\ \alpha_7^* \end{bmatrix}_{7\times 1} + e_{208\times 1} \ . \tag{4.21}$$

In equation (4.21) we have the following:

$${}^*y^I = y^I - f^I(\boldsymbol{\alpha}^{(0)}, t) \text{ and } {}^*y^P = y^P - f^P(\boldsymbol{\alpha}^{(0)}, t) \tag{4.22}$$

and

$$\alpha_j^* = \alpha_j - \alpha_j^{(0)}. \tag{4.23}$$

In equations 4.22 and 4.23 $\alpha^{(0)}$ is the initial estimate for α. This initial estimate will be replaced with updated estimates until convergence. The complete error vector has the following distribution:

$$\mathbf{e} \sim N[\mathbf{0}, \mathbf{V}(\boldsymbol{\theta})] \text{ where } \mathbf{V}(\boldsymbol{\theta}) = \sum_q \theta_q \mathbf{V}_q \,. \tag{4.24}$$

The problem is to correctly specify the $\mathbf{V}_q$ to find the "best" variance structure. For the purposes of this example we will examine several variance structures.

4.3.1 Model 1: Complete Independence

For this model we will assume that the error structure is given by independent heterogeneous variances of phosphate and insulin. There will be no correlation between insulin and phosphate. For this model we only need two variances components. The $\mathbf{V}_q$ are given by:

$$\mathbf{V}_1 = \begin{bmatrix} \mathbf{I}_{104} & \mathbf{0}_{104} \\ \mathbf{0}_{104} & \mathbf{0}_{104} \end{bmatrix} \text{ and } \mathbf{V}_2 = \begin{bmatrix} \mathbf{0}_{104} & \mathbf{0}_{104} \\ \mathbf{0}_{104} & \mathbf{I}_{104} \end{bmatrix} . \tag{4.25}$$

Here $\mathbf{I}_{104}$ is a 104×104 identity matrix and $\mathbf{0}_{104}$ is a 104×104 matrix of zeros. The -2 log likelihood and AIC for this model are listed in Table 4.6 (page 76). AIC is Akaike's Information Criterion (Akaike, 1973; Jones, 1993). It is calculated as:

$$\text{AIC} = -2 \ln \text{likelihood} + 2(\#\text{parameters in model}). \tag{4.26}$$

A "rule of thumb" for using AIC is to choose all models within 2 of the lowest AIC as possible "best" models, and from this list of models choose the most parsimonious model (Jones, 1993; Duong, 1984) .

4.3.2 Model 2: Independent within Phosphate and Insulin with Correlation between Phosphate and Insulin

For this model we need 3 variance components. The corresponding variance design matrices are:

$$\mathbf{V}_1 = \begin{bmatrix} \mathbf{I}_{104} & \mathbf{0}_{104} \\ \mathbf{0}_{104} & \mathbf{0}_{104} \end{bmatrix}, \mathbf{V}_2 = \begin{bmatrix} \mathbf{0}_{104} & \mathbf{0}_{104} \\ \mathbf{0}_{104} & \mathbf{I}_{104} \end{bmatrix}, \text{ and } \mathbf{V}_3 = \begin{bmatrix} \mathbf{0}_{104} & \mathbf{I}_{104} \\ \mathbf{I}_{104} & \mathbf{0}_{104} \end{bmatrix}. \tag{4.27}$$

4.3.3 Model 3: Heterogeneous Compound Symmetric

For the heterogeneous compound symmetric variance structure we need the following variance design matrices:

WITHIN

$$\mathbf{V}_1 = \begin{bmatrix} \mathbf{I}_{104} & \mathbf{0}_{104} \\ \mathbf{0}_{104} & \mathbf{0}_{104} \end{bmatrix} \quad \mathbf{V}_2 = \begin{bmatrix} \mathbf{0}_{104} & \mathbf{0}_{104} \\ \mathbf{0}_{104} & \mathbf{I}_{104} \end{bmatrix}$$

BETWEEN

$$\mathbf{V}_3 = \begin{bmatrix} \mathbf{I}_{13} \otimes \mathbf{J}_8 & \mathbf{0}_{104} \\ \mathbf{0}_{104} & \mathbf{0}_{104} \end{bmatrix} \quad \mathbf{V}_4 = \begin{bmatrix} \mathbf{0}_{104} & \mathbf{0}_{104} \\ \mathbf{0}_{104} & \mathbf{I}_{13} \otimes \mathbf{J}_8 \end{bmatrix}. \tag{4.28}$$

$\mathbf{J}_8$ is an 8×8 matrix of ones and $\otimes$ is the Kroenecker product. Looking ahead to Table 4.6, it can be seen that by using the compound symmetric variance structure, the AIC is reduced.

4.3.4 Model 4: Multivariate Compound Symmetric

If we wish to include a between and within cross correlation between insulin and phosphate, we need to add two parameters. The variance design matrices are:

$$\mathbf{V}_1 = \begin{bmatrix} \mathbf{I}_{104} & \mathbf{0}_{104} \\ \mathbf{0}_{104} & \mathbf{0}_{104} \end{bmatrix} \quad \mathbf{V}_2 = \begin{bmatrix} \mathbf{0}_{104} & \mathbf{0}_{104} \\ \mathbf{0}_{104} & \mathbf{I}_{104} \end{bmatrix} \quad \mathbf{V}_3 = \begin{bmatrix} \mathbf{0}_{104} & \mathbf{I}_{104} \\ \mathbf{I}_{104} & \mathbf{0}_{104} \end{bmatrix}$$

and

$$\mathbf{V}_4 = \begin{bmatrix} \mathbf{I}_{13} \otimes \mathbf{J}_8 & \mathbf{0}_{104} \\ \mathbf{0}_{104} & \mathbf{0}_{104} \end{bmatrix} \quad \mathbf{V}_5 = \begin{bmatrix} \mathbf{0}_{104} & \mathbf{0}_{104} \\ \mathbf{0}_{104} & \mathbf{I}_{13} \otimes \mathbf{J}_8 \end{bmatrix} \quad \mathbf{V}_6 = \begin{bmatrix} \mathbf{0}_{104} & \mathbf{I}_{13} \otimes \mathbf{J}_8 \\ \mathbf{I}_{13} \otimes \mathbf{J}_8 & \mathbf{0}_{104} \end{bmatrix}. \quad (4.29)$$

From the results shown in Table 4.6, it does not appear that accounting for the cross correlation ($\mathbf{V}_3$ and $\mathbf{V}_6$) reduces the AIC.

4.3.5 Model 5: Semi-banded (2 bands)

For the semi-banded error structure we need to define the following matrices:

$$\mathbf{S}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{and } \mathbf{S}_2 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}. \quad (4.30)$$

The variance design matrices are as follows:

$$\mathbf{V}_1 = \begin{bmatrix} \mathbf{I}_{13} \otimes \mathbf{S}_1 & \mathbf{0}_{104} \\ \mathbf{0}_{104} & \mathbf{0}_{104} \end{bmatrix} \qquad \mathbf{V}_2 = \begin{bmatrix} \mathbf{I}_{13} \otimes \mathbf{S}_2 & \mathbf{0}_{104} \\ \mathbf{0}_{104} & \mathbf{0}_{104} \end{bmatrix} \tag{4.31}$$

$$\mathbf{V}_3 = \begin{bmatrix} \mathbf{0}_{104} & \mathbf{0}_{104} \\ \mathbf{0}_{104} & \mathbf{I}_{13} \otimes \mathbf{S}_1 \end{bmatrix} \qquad \mathbf{V}_4 = \begin{bmatrix} \mathbf{0}_{104} & \mathbf{0}_{104} \\ \mathbf{0}_{104} & \mathbf{I}_{13} \otimes \mathbf{S}_2 \end{bmatrix} \tag{4.32}$$

$$\mathbf{V}_5 = \begin{bmatrix} \mathbf{0}_{104} & \mathbf{I}_{13} \otimes \mathbf{S}_1 \\ \mathbf{I}_{13} \otimes \mathbf{S}_1 & \mathbf{0}_{104} \end{bmatrix} \quad \mathbf{V}_6 = \begin{bmatrix} \mathbf{0}_{104} & \mathbf{I}_{13} \otimes \mathbf{S}_2 \\ \mathbf{I}_{13} \otimes \mathbf{S}_2 & \mathbf{0}_{104} \end{bmatrix}. \tag{4.33}$$

To get an idea of what this variance structure looks like, we can examine what an individual's variance/covariance matrix looks like. If the 8 individual insulin values are stacked on top of the 8 individual phosphate values, the variance/covariance matrix will look like the following:

INSULIN

$$\left[\begin{array}{cccccccc:cccccccc}
\theta_1 & \theta_2 & 0 & 0 & 0 & 0 & 0 & 0 & \theta_5 & \theta_6 & 0 & 0 & 0 & 0 & 0 & 0 \\
\theta_2 & \theta_1 & \theta_2 & 0 & 0 & 0 & 0 & 0 & \theta_6 & \theta_5 & \theta_6 & 0 & 0 & 0 & 0 & 0 \\
0 & \theta_2 & \theta_1 & \theta_2 & 0 & 0 & 0 & 0 & 0 & \theta_6 & \theta_5 & \theta_6 & 0 & 0 & 0 & 0 \\
0 & 0 & \theta_2 & \theta_1 & \theta_2 & 0 & 0 & 0 & 0 & 0 & \theta_6 & \theta_5 & \theta_6 & 0 & 0 & 0 \\
0 & 0 & 0 & \theta_2 & \theta_1 & \theta_2 & 0 & 0 & 0 & 0 & 0 & \theta_6 & \theta_5 & \theta_6 & 0 & 0 \\
0 & 0 & 0 & 0 & \theta_2 & \theta_1 & \theta_2 & 0 & 0 & 0 & 0 & 0 & \theta_6 & \theta_5 & \theta_6 & 0 \\
0 & 0 & 0 & 0 & 0 & \theta_2 & \theta_1 & \theta_2 & 0 & 0 & 0 & 0 & 0 & \theta_6 & \theta_5 & \theta_6 \\
0 & 0 & 0 & 0 & 0 & 0 & \theta_2 & \theta_1 & 0 & 0 & 0 & 0 & 0 & 0 & \theta_6 & \theta_5 \\
\hdashline
\theta_5 & \theta_6 & 0 & 0 & 0 & 0 & 0 & 0 & \theta_3 & \theta_4 & 0 & 0 & 0 & 0 & 0 & 0 \\
\theta_6 & \theta_5 & \theta_6 & 0 & 0 & 0 & 0 & 0 & \theta_4 & \theta_3 & \theta_4 & 0 & 0 & 0 & 0 & 0 \\
0 & \theta_6 & \theta_5 & \theta_6 & 0 & 0 & 0 & 0 & 0 & \theta_4 & \theta_3 & \theta_4 & 0 & 0 & 0 & 0 \\
0 & 0 & \theta_6 & \theta_5 & \theta_6 & 0 & 0 & 0 & 0 & 0 & \theta_4 & \theta_3 & \theta_4 & 0 & 0 & 0 \\
0 & 0 & 0 & \theta_6 & \theta_5 & \theta_6 & 0 & 0 & 0 & 0 & 0 & \theta_4 & \theta_3 & \theta_4 & 0 & 0 \\
0 & 0 & 0 & 0 & \theta_6 & \theta_5 & \theta_6 & 0 & 0 & 0 & 0 & 0 & \theta_4 & \theta_3 & \theta_4 & 0 \\
0 & 0 & 0 & 0 & 0 & \theta_6 & \theta_5 & \theta_6 & 0 & 0 & 0 & 0 & 0 & \theta_4 & \theta_3 & \theta_4 \\
0 & 0 & 0 & 0 & 0 & 0 & \theta_6 & \theta_5 & 0 & 0 & 0 & 0 & 0 & 0 & \theta_4 & \theta_3
\end{array}\right]$$

PHOSPHATE

4.3.6 Model 6-11: More Semi-Banded

In a similar fashion to section 4.3.5 we can add "bands" to the covariance structure until we have a multivariate "fully" banded error structure. For the purposes of brevity, all semi-banded error structures will not be listed. The results using these models are presented in Table 4.6. It should be noted that the multivariate fully banded error structure failed to converge. This failure was the result of the variance/covariance matrix going negative definite. This is a drawback to using the Hocking algorithm. There is no built in prevention of negative definite matrices.

4.3.7 Model 12: Almost Totally Banded

In an attempt to obtain something close to the "fully" banded error structure the 7th band was omitted and all other bands were included. This model was not significantly better than model (9).

4.3.8 Model 13: Heterogeneous Totally Banded

The meaning of a banded error structure for the covariance between insulin and phosphate is questionable. For this reason a model that allows for a banded error structure for insulin and a different banded error structure for phosphate was attempted. The model did not allow for any covariance between insulin and phosphate. As can be seen in Table 4.6 this model beats all of the previous models. The only one that comes close is model (9).

4.3.9 Model 14: Heterogeneous Totally Banded with Insulin/Phosphate Covariance

This model is very similar to model (13). The only difference is that we allow for insulin and phosphate values taken at the same time to be correlated. This model gives the lowest AIC of all the models attempted.

4.3.10 Model 15: Stochastic Parameter Model.

In this model each parameter of (4.19) and (4.20) was allowed to vary stochastically across individuals. This model had the most parameters of any of the models attempted. It also had the lowest -2 log likelihood.

Table 4.6

Results of fitting different variance structures to the data in Table 4.5.

Variance Structure	# of Parameters	-2 ln Likelihood	AIC	time lag +/- 1 S.E.
1. Comp. Independent	7+2=9	560.48	578.48	40.61 +/- 26.53
2. Semi-banded (1 band)	7+3=10	558.83	578.83	40.44 +/- 26.50
3. Hetro. Comp. Symm.	7+4=11	491.32	513.32	40.61 +/- 20.66
4. Mult. Comp. Symm.	7+6=13	487.76	513.76	40.29 +/- 20.55
5. Semi-banded (2 bands)	7+6=13	490.64	516.64	46.61 +/- 31.98
6. Semi-banded (3 bands)	7+9=16	482.55	514.55	49.21 +/- 33.43
7. Semi-banded (4 bands)	7+12=19	468.36	506.36	44.91 +/- 29.59
8. Semi-banded (5 bands)	7+15=22	465.69	509.69	42.35 +/- 26.24
9. Semi-banded (6 bands)	7+18=25	446.88	496.88	50.21 +/- 22.56
10. Semi-banded (7 bands)	7+21=28	444.68	500.68	53.80 +/- 27.09
11. Totally banded (8 bands)	7+24=31	*	*	*
12. Semi-banded (1,2,3,4,5,6,8)	7+21=28	444.23	500.23	50.13 +/- 22.71
13. Tot. Banded no Corr.	7+16=23	446.89	492.89	47.50 +/- 23.90
14. Tot. Banded with Corr.	7+17=24	440.89	488.89	48.01 +/- 21.66
15. Stochastic Parameter	7+31=38	433.19	509.19	37.28 +/- 16.49

*NOTE: * Could not converge.*

4.3.11 Estimating the Time Lag Between Maximum Insulin and Minimum Phosphate

The time lag between maximum insulin and minimum phosphate values was estimated as a nonlinear function of the fixed effects in the model. Taking the derivative of (4.19) with respect to "t" and equating to zero gives the maximum insulin value occurring at:

$$t^{I}_{\max} = \alpha_2 \cdot \alpha_3 . \tag{4.34}$$

Taking the derivative of (4.20) and equating the result to zero and solving for "t" gives the minimum phosphate value occurring at:

$$t^{P}_{\min} = \sqrt{\alpha_7/\alpha_5} - \alpha_6 . \tag{4.35}$$

Now the nonlinear function of the parameters that we wish to estimate is given by:

$$g(\boldsymbol{\alpha}) = t^{P}_{\min} - t^{I}_{\max} = (\sqrt{\alpha_7/\alpha_5} - \alpha_6) - (\alpha_2 \cdot \alpha_3). \tag{4.36}$$

Again following the work of Zerbe et al. (1994), the nonlinear function $g(\boldsymbol{\alpha})$ can be expanded in a Taylor series about $\boldsymbol{\alpha}$. We therefore have:

$$g(\hat{\boldsymbol{\alpha}}) \approx g(\boldsymbol{\alpha}) + [\partial g/\partial \boldsymbol{\alpha}]'|_{\alpha=\hat{\alpha}}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}). \tag{4.37}$$

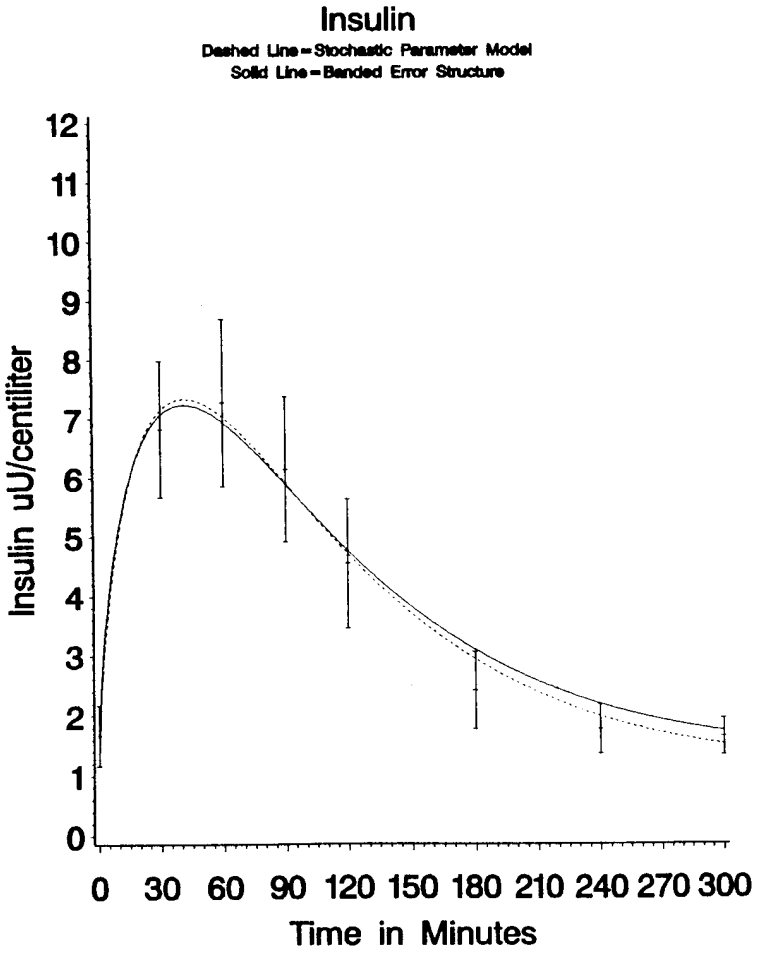
Thus $g(\hat{\boldsymbol{\alpha}})$ is asymptotically normally distributed with:

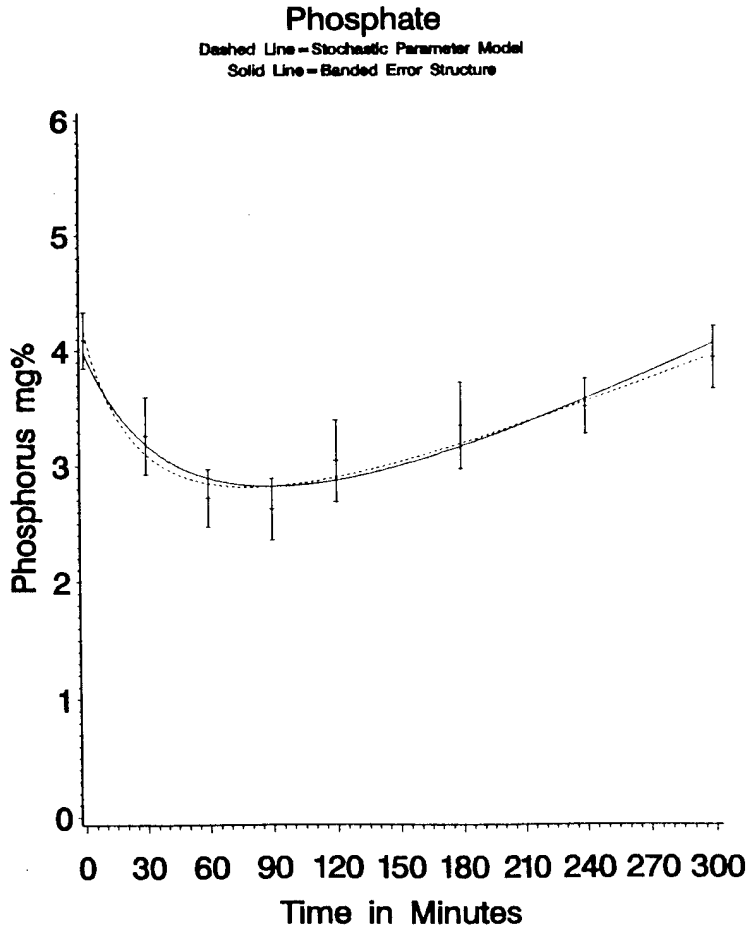
$$E[g(\hat{\boldsymbol{\alpha}})] \approx g(\boldsymbol{\alpha}) \text{ and } \mathbf{V}[g(\alpha)] \approx [\partial g/\partial \boldsymbol{\alpha}]'\hat{\mathbf{V}}(\boldsymbol{\alpha})[\partial g/\partial \boldsymbol{\alpha}]. \tag{4.38}$$

In (4.38) we have $\hat{\mathbf{V}}(\boldsymbol{\alpha}) \approx \left(\mathbf{H}'\mathbf{V}^{-1}\mathbf{H}\right)^{-1}$. Here $\mathbf{H}$ is the design matrix from the last iteration of the estimation process in model (4.21) . The estimates of the time lag along with the standard errors are listed in Table 4.6. Table 4.6 also shows the importance of choosing the correct variance structure. Choosing the wrong error structure can result in inconsistent estimates and standard errors (van Houwelingen, 1988). Looking at Figure 4.2, it can be seen that the function for phosphate is rather flat around the minimum. As a result, slight changes in the estimates of the function will lead to large changes in the estimate of when the minimum occurs. The stochastic parameter model has the lowest -2 log likelihood, but its large number of parameters increase the AIC dramatically. Even though by the AIC, model (14) is "better" than model (15), there are several advantages to choosing model (15). First, it seems very intuitive to fit the nonlinear equations (4.19) and (4.20) to each individual and then summarize across individuals. This is essentially what the stochastic parameter model is doing. Secondly, looking at the plots of model (14) and model (15), it appears that model (15) has a slightly better fit. Figure 4.2 shows models (14) and (15) along with the average value for the insulin and phosphate data. Model (15) consistently comes closer to the average value for both insulin and phosphate. Thirdly, concerning the estimated time lag between minimum phosphate and maximum insulin, model (15) has the lowest standard error of all the competing models. Appendix C contains the MATLAB computer program used to run model (14). Appendix D is the output from the MATLAB program.

Figure 4.2

Models 14 and 15 plotted against each other and average values. +/- 2 standard deviations of the mean are also plotted.





CHAPTER V

COMMENTS

5.1 Warnings

This thesis extends the current methods of analyzing nonlinear longitudinal data to the case of multi-response data. Handling the multiple responses is essentially accomplished by stacking the models in a clever fashion and then keeping track of which observations contribute to the different variance components. Besides this element, the MNLMEM is essentially a nonlinear model of the type discussed by Lindstrom and Bates (1990), Hirst et al. (1991), and Young et al. (1992). As a result the MNLMEM suffers from all of the pitfalls of the previously mentioned methods. These problems are compounded by the fact that MNLMEM deals with multi-response data, therefore, generally speaking the MNLMEM has more fixed effects, random effects, and variance components to estimate. Three pitfalls particularly salient in the practical application of MNLMEM will be discussed next.

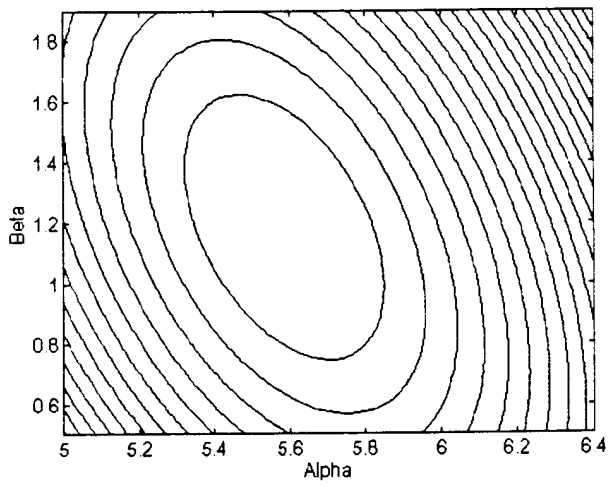
First, the iterative procedure may converge to the wrong answer or not converge at all. Second, there may be ill-conditioning in parameter estimation. Thirdly, the application of asymptotic theory may not be appropriate.

Figure 5.1 shows the -2 log likelihood contour plot for the simple linear regression model $y = \alpha + \beta x + e$. As can be seen, the likelihood surface has an elliptical contour. Finding the minimum of such a surface is computationally simple. In fact, one can express the result in closed form. This simplicity is compromised when we start dealing with a nonlinear model. In Figure 5.2 the least squares contours for the nonlinear

model $y = \alpha \exp\{-\beta x\} + e$ are plotted (Seber and Wild, 1989). This surface will be very similar to the -2 log likelihood surface, differing only by a constant.

Figure 5.1

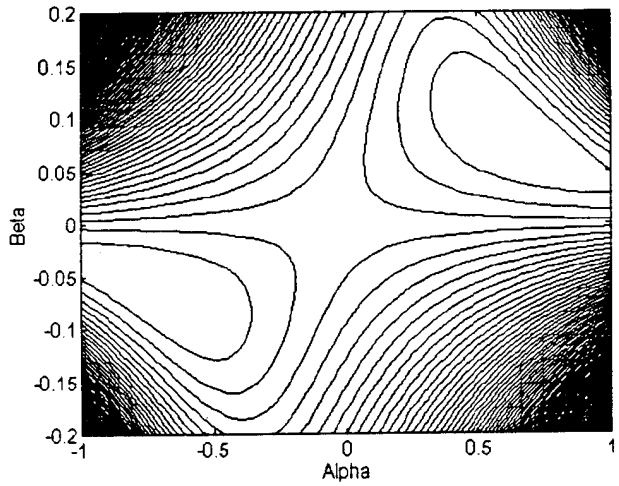
Likelihood surface for a simple linear model.



As can be seen in Figure 5.2 there are two local minima on the likelihood surface, only one of which is also the global minimum. If the wrong starting point is chosen, the iterative process will converge to the wrong answer.

Figure 5.2

Least squares surface of a nonlinear model.



Contour plots can be helpful when there are only two parameters in the model. Since many models have more than two parameters, it would be possible to look at all pairs of parameters. However, this soon becomes cumbersome and difficult to carry out in practice. In general, nonlinear models produce likelihood surfaces that have local maximums and have varying degrees of curvature.

In the linear model $\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, ill-conditioning occurs when the design matrix $\mathbf{X}$ is not of full rank. This often occurs in analysis of variance models as opposed to regression models. When $\mathbf{X}$ is not full rank, there are an infinite number of solutions to the normal equations. This problem is usually handled by implementing restrictions such as $\beta_i = 0$ or $\Sigma_i \beta_i = 0$ or looking at linear combinations $\mathbf{t}'\boldsymbol{\beta}$ that are invariant to the choice of $\boldsymbol{\beta}$ (Myers and Milton, 1991). In regression models that use continuous data, $\mathbf{X}$ is almost always of full rank. However, the columns of $\mathbf{X}$ are sometimes close to being linearly dependent. This leads to a problem called multicollinearity. The likelihood function for a linear model that has multicollinearity will have likelihood contours that are long skinny ellipses. This translates to parameter estimates having large variances. Multicollinearity also makes the accurate computation of $(\mathbf{X}'\mathbf{X})^{-1}$ exceedingly more difficult. Some computer algorithms will warn the user that impending results may be inaccurate. By linearizing the nonlinear model, the design matrix [$\mathbf{H}$, from (3.37) for example] is changing with each iteration. Thus, with each iteration it is possible to run into multicollinearity problems. Another problem that can lead to ill-conditioning in nonlinear models is having a nonlinear model that fits the data well for many different values of parameters. This can lead to a flat likelihood surface in the direction of that

parameter. Sometimes a simple reparameterization of the nonlinear model can fix the problem and yield more precise estimates of the parameters. The likelihood surfaces for single-response independent nonlinear models are often banana shaped and these models don't have the complicated error structure associated with longitudinal data.

Consider the following univariate nonlinear fixed effects model:

$$\mathbf{y} = \mathbf{f}(\boldsymbol{\beta}, \mathbf{x}) + e, \text{ where } e \sim N(0, \sigma^2\mathbf{I}). \tag{5.1}$$

Typically to obtain the maximum likelihood estimates this model is "linearized" via Taylor series expansion. The Taylor series expansion can be thought of as a computational crutch used to find the maximum of the likelihood surface. Once the algorithm converges the usual asymptotic property that is used to make inference is:

$$\hat{\boldsymbol{\alpha}} \sim N[\mathbf{0}, \sigma^2(\mathbf{H}'\mathbf{H})^{-1}] \text{ where } \mathbf{H} = \partial\mathbf{f}/\partial\boldsymbol{\alpha}'\big|_{\alpha=\hat{\alpha}}. \tag{5.2}$$

It is often the case that (5.2) yields misleading results because the linearized likelihood surface does not approximate the true likelihood surface adequately (Seber and Wild, 1989). In the nonlinear regression setting this problem is known as curvature. By looking at the second derivatives of the nonlinear function **f** it is possible to measure the degree of bending and twisting in the surface **f**; and the amount of curvature induced by the choice of parameters. The MNLMEM models with additive errors suffer from the same effects of curvature. However, nonlinear mixed effects models with random effects in the nonlinear part of the model suffer from an additional level of approximation. For

example, consider the model (2.62) by Hirst et al. (1991). They suggest using the following to make inference about the fixed effects:

$$\hat{\boldsymbol{\alpha}} \sim N(\boldsymbol{\alpha}, \sum_{i=1}^{M} \mathbf{H}_i'[\mathbf{Z}_i \mathbf{D} \mathbf{Z}_i' + \sigma^2 \mathbf{I}]^{-1} \mathbf{H}_i). \tag{5.3}$$

The variance in (5.3) not only depends on linearization about the fixed effects ($\mathbf{H}_i$) but also the linearization about the random effects ($\mathbf{Z}_i$). This is a consequence of approximating the conditional distribution $\mathbf{y}_i \mid \mathbf{b}_i$. By linearizing with respect to the random effects, we are essentially approximating the likelihood surface around the given estimates with a likelihood surface whose maximum we know how to find. Aside from a simulation study by Pineiro and Bates (1995), little has been done to show how well this approximation works. Pineiro and Bates (1995) concluded that for their example the approximation works well.

5.2 Future Research

The distributional properties of the estimates of nonlinear mixed effects models have not clearly been established. Currently, most authors treat the estimates as maximum likelihood or restricted maximum likelihood estimates. A way of describing how congruent the approximate likelihood surface is with the true likelihood surface would be an area for future research. This would probably involve trying to tackle the integral (2.52) using state of the art numerical analysis techniques. As stated earlier the MNLMEM suffers from these same problems and there is more potential for poor estimation since there are more fixed effects, random effects, and variance components

associated with a MNLMEM model. In addition, the little work that has been accomplished on the properties of the estimates focuses mainly on the fixed effects. An advantage of the MNLMEM method is that functions of the variance components frequently have interpretations such as correlations or regression coefficients between different parameters. Therefore, a better understanding of the properties of the variance components could be a fruitful area for future research. The model with additive errors stands on more solid theoretical ground than the model with nonlinear random effects. However, it is the stochastic parameter model of section 3.1 that seems to be able to answer many intriguing questions. For example, suppose a researcher has bivariate data and each outcome can be modeled with a Michaelis-Menten equation. Then, using the stochastic parameter model it would be possible to determine the association between the two different Michaelis-Menten constants.

5.3 Summary

This thesis extends the current methodology for analyzing nonlinear longitudinal data by allowing for the modeling of multiple responses. Currently, researchers often ignore the correlation of multiple responses taken on the same subject and perform separate univariate analysis on each response. Two multi-response nonlinear mixed effects models (MNLMEM) were proposed. The models both revolved around using a Taylor series expansion to reduce the nonlinear model to a model with known techniques of parameter estimation. In the first MNLMEM, random effects were allowed into the nonlinear part of the model. The advantage of this model is that it allows the researcher to answer questions about how the parameters of the different curves relate to each other.

The disadvantage of this model is that the asymptotic properties of the parameter estimates have not been well established. In the second MNLMEM, the error structure is additive and therefore the asymptotic theory is better understood. Using the additive error structure it is possible to model the multi-response analogs of the univariate compound symmetric, banded, and arbitrary error structures. I feel that the major contribution of this thesis is that it opens up a whole new area for potential research in the medical field. Just how do you describe how two or more nonlinear functions behave over time or space. For example, when I take my son to the doctor they always plot his height and weight on a growth chart. How are height and weight in children related over time? Does weight increase at the same rate as height? If a child is short to start with will his or her weight increase faster or slower than a taller child? Multi-response data is often collected but seldom analyzed multivariately. This is largely due to the fact that the methods involved are difficult to implement, often require balanced data, or have yet to be developed. Also, adequate computing power has become cheaply available only within recent years. This thesis will allow researchers to consider modeling their nonlinear multi-response longitudinal data multivariately.

REFERENCES

Airy, G.B. (1861). *On the Algebraical and Numerical Theory of Errors of Observations and the Combinations of Observations.* MacMillan, London.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *Second International Symposium on Information Theory*, (B.N. Petrov and F. Csaki, Eds.), Budapest: Akademia Kaido, 267-81.

Bates, D.M. and Watts, D.G. (1981). A relative offset orthogonality convergence criterion for nonlinear least squares. *Technometrics* **23**: 179-183.

Bates, D.M. and Watts, D.G. (1987). A generalized Gauss-Newton procedure for multi-response parameter estimation. *SIAM J. Sci. and Stat Comput.* **8**: 49-55

Berkey, C.S. (1982). Bayesian approach for a nonlinear growth model. *Biometrics* **38**: 953-961.

Box, G.E.P. (1950). Problems in the analysis of growth and wear curves. *Biometrics* **6**: 362-389.

Casella G. and Berger R.L. (1990). *Statistical Inference*. Brooks/Cole Publishing Company, Belmont, CA.

Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B* **39**: 1-38.

Draper, N.R. and Smith, H. (1981). *Applied Regression Analysis*. John Wiley and Sons, New York.

Duong, Q.P. (1984). On the choice of the order of autoregressive models: a ranking and selection approach. *Journal of Time Series Analysis* **5**:145-158.

Fisher, R.A. (1925). *Statistical Methods for Research Workers*, 1st edn. Oliver & Boyd, Edinburgh and London.

Grizzle, J.E. and Allen, D.M. (1969). Analysis of growth and dose response curves, *Biometrics* **25**:357-382.

Hartley, H.O. and Rao, J.N.K. (1967). Maximum likelihood estimation for the mixed analysis of variance model. *Biometrika* **54**: 93-108.

Harville, D.A. (1976). Extension of the Gauss-Markov theorem to include the estimation of random effects. *Annals of Statistics* **4**: 384-395.

Harville, D.A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association* **72**: 320-340.

Henderson, C.R. (1953). Estimation of variance and covariance components. *Biometrics* **9**: 226-252.

Hirst, K., Zerbe, G.O., Boyle, D.W., and Wilkening, R.B. (1991). On nonlinear random effects models for repeated measurements. *Communications In Statistics B, Simulation and Computation* **20**:463-478.

Hocking, R.R. (1985). *The Analysis of Linear Models.* Brooks/Cole, Monterey, CA.

James, W. and Stein, C. (1961). Estimation with quadratic loss. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* **1**: 361-380. Berkley, California: University of California Press.

Jennrich, R.I. and Schluchter, M.D. (1986). Unbalanced repeated-measures models with structured covariance matrices. *Biometrics* **42**:805-820.

Jones, R.H. (1990). Serial correlation or random subject effect? *Communications In Statistics B, Simulation and Computation* **19**: 1105-1123.

Jones, R.H. and Boadi-Boateng, F. (1991). Unequally spaced longitudinal data with AR(1) serial correlation. *Biometrics* **47**: 161-175.

Jones, R.H. (1993). *Longitudinal Data with Serial Correlation: A State Space Approach.* Chapman and Hall, New York.

Khatri, C.G. (1966). A note on a MANOVA model applied to problems in growth curves. *Annals of the Institute of Statistical Mathematics* **18**: 75-86.

Laird, N.M. and Ware, J.H. (1982). Random effects models for longitudinal data. *Biometrics* **38:**963-974.

Liang, K.Y. and Zeger, S.L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**: 13-22.

Lindley, D.V. and Smith, A.F.M. (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society, Series B* **34**: 1-42.

Lindstrom, M.J. and Bates, D.M. (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics* **46**:673-687.

Lynch, A., Marlar, R., Murphy, J., Davilla, G., Santos, M., Rutledge, J., and Emlen, W. (1994). Antiphospholipid Antibodies in Predicting Adverse Pregnancy Outcome. *Annals of Internal Medicine* **120 #6**: 470-475.

Magnus, J.R. and Neudecker, H. (1988). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley & Sons, New York.

McClatchey, M.W. (1993). A sequence of methods for estimation of human energy expenditure by doubly labeled water techniques. *Ph.D. Thesis.* University of Colorado Health Sciences Center.

McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models.* Chapman and Hall, New York.

Mickey, R.H., Shema, S.J., Vacek, P.M., and Bell, D.Y. (1994). Analysis of multiple outcome variables measured longitudinally. *Computational Statistics and Data Analysis* **17**: 17-33.

Myers, R.H. and Milton, J.S. (1991). *A First Course in the Theory of Linear Statistical Models.* PWS-Kent Publishing Comany, Boston.

Patterson, H.D. and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika* **58**: 545-554.

Pinheiro, J.C. and Bates, D.M. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics* **4**: 12-35.

Pothoff, R.F. and Roy, S.N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika* **51**: 313-326.

Prentice A.M. (1990). The doubly-labeled water method: Technical recommendations for use in humans, a consensus report by the IDECG working group. Conference held September, Claire College, Cambridge.

Racine-Poon, A. (1985). A Bayesian approach to nonlinear random effects models. *Biometrics* **41**: 1015-1023.

Rao, C.R. (1958). Some statistical methods for comparison of growth curves. *Biometrics* **14**: 1-17.

Rao, C.R. (1959). Some problems involving linear hypotheses in multivariate analysis. *Biometrika* **46**: 49-58.

Rao, C.R. (1965). The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves. *Biometrika* **52**: 447-457.

Ravussin, E., Harper, I.T., Rising, R., and Bogardus, C. (1991) Energy expenditure by doubly labeled water: validation in lean and obese subjects. *American Journal of Physiology* **261**, (Endocrinol. Metab. 24), E402-E409.

Roy, S.N. (1957). *Some Aspects of Multivariate Analysis.* New York, John Wiley and Sons.

SAS Institute Inc., Cary, NC.

Schoeller, D. A. (1988). Measurement of energy expenditure in free-living humans by using doubly labeled water, *The Journal of Nutrition* **118**: 1278-1289.

Searle, S.R., Casella, G., and McCulloch, C.E. (1992). *Variance Components.* John Wiley and Sons, New York.

Seber, G.A.F. and Wild, C.J. (1989). *Nonlinear regression.* John Wiley and Sons, New York.

Sheiner, L.B. and Beal, S.L. (1980) Evaluation of methods for estimating population pharmacokinetic parameters. I. Michaelis-Menten model: routine clinical pharmacokinetic data. *J. Pharmacokin. Biopharm*, **8:** 553-571.

Thompson, W.A. (1962). The problem of negative estimates of variance components. *Annals of Mathematical Statistics* **33**: 273-289.

van Houwelingen, J.C. (1988). Use and abuse of variance models in regression. *Biometrics* **44**: 1073-1081.

Vonesh, E.F. (1992). Non-linear models for the analysis of longitudinal data. *Statistics in Medicine* **11**: 1929-1954.

Vonesh, E.F. and Carter, R.L. (1992). Mixed-effects nonlinear regression for unbalanced repeated measures. *Biometrics* **48**: 1-17.

Wishart, J. (1938). Growth-rate determinations in nutrition studies with the bacon pig, and their analysis. *Biometrica* **30**: 16-28.

Wu, C.F. (1983). On the convergence properties of the EM algorithm. *Annals of Statistics* **11**: 95-103.

Young, D.A., Zerbe, G.O., and Hay, W.W. (1992). Application of a non-linear random-effects model for comparing x-intercepts of linear regression lines. *Statistics in Medicine* **11**: 2039-2040.

Zeger, S.L., Liang, K.Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* **42**: 121-130.

Zeger, S.L., Liang, K.Y., and Albert, P.S. (1988). Models for longitudinal data: a generalized estimating equation approach. *Biometrics* **44**: 1049-1060.

Zeger, S.L., Liang, K.Y. (1992). An overview of methods for the analysis of longitudinal data. *Statistics in Medicine* **11**: 1825-1839.

Zerbe, G.O., Wu, M.C., and Zucker, D.M. (1994). Studying the relationship between change and initial value in longitudinal studies. *Statistics in Medicine* **13:** 759-768.

Zucker, D.M., Zerbe, G.O., and Wu, M.C. (In press). Inference for the association between coefficients in a multivariate growth curve model. *Biometrics*.

APPENDIX A

SAS/IML Program Implementing MNLMEM on the Doubly Labeled Water Data Set

```
/*-------------------------------------------------------------
    MULTIVARIATE NONLINEAR MODEL WITH RANDOM EFFECTS
        JIM RUTLEDGES VERSION OF MARCH 19, 1995

  This program is based on the model, notation, and
  EM algorithm discussed by Nan Laird and James Ware
  (1982). Random-Effects Models for Longitudinal Data.
  Biometrics 38, 963-974.  The program is a modified
  version of a SAS/IML program written by Gary O. Zerbe,
  Ph.D., University of Colorado Health Sciences Center.

      PROGRAM:  dblwat4d.iml
         DATE:  4 November 1994
  DESCRIPTION:  This program is used on M. McClatchey's doubly
                labeled water data.  12 subjects with missing data.
                The within subject errors are not independent.
                The model has not been reparameterized.
                The bi's are stored in a matrix called b.
                The user may choose Lindstrom and Bates or
                Sheiner and Beale.  In addition either ML or REML
                may be selected.

-------------------------------------------------------------*/


options nocenter pagesize=59 linesize=100;

proc iml worksize=600; reset log nocenter fw=15;

/*   INITIALIZATION   */

start init;

rawdata={
  1     0.30   1  3.48854    3.35849,
  1     0.45   2  3.32859    3.26164,
  1     0.80   3  3.23474    3.16630,
  1     1.80   4  2.97045    2.93978,
  1     2.80   5  2.74969    2.80495,
  1     3.80   6  2.52143    2.62949,
  1     5.80   7  2.11144    2.25653,
  1     7.80   8  1.76980    1.96274,
  2     0.30   1  4.84651    4.64355,
  2     0.45   2  4.57692    4.39669,
  2     0.80   3  4.43260    4.30189,
  2     1.80   4  4.07958    3.96784,
  2     2.80   5  3.63036    3.65238,
  2     3.80   6  3.25812    3.35920,
  2     5.80   7  2.60030    2.76285,
  2     7.80   8  2.10185    2.39742,
  3     0.30   1  3.20667    3.11547,
  3     0.45   2  3.12248    3.04462,
  3     0.80   3  3.04104    2.99848,
  3     1.80   4  2.67324    2.59952,
  3     2.80   5  2.37099    2.41369,
  3     3.80   6  2.10948    2.19710,
  3     4.80   7  1.90107    -999,
  3     5.80   8  1.69543    1.84347,
  3     6.80   9  1.50291    -999,
  3     7.80  10  1.36215    1.53534,
  4     0.45   1  3.81704    3.63285,
  4     0.80   2  3.67215    3.54392,
  4     1.80   3  3.20852    3.15908,
```

4	2.80	4	2.72821	2.72974,
4	3.80	5	2.36556	2.43855,
4	4.80	6	2.03101	-999,
4	5.80	7	1.76670	1.91005,
5	0.30	1	3.98354	3.83665,
5	0.45	2	3.89723	3.76665,
5	0.80	3	3.80345	3.61770,
5	1.80	4	3.40261	3.38561,
5	2.80	5	3.08061	3.08211,
5	3.80	6	2.84823	2.87339,
5	4.80	7	2.61171	2.70545,
5	5.80	8	2.41170	2.53691,
5	6.80	9	2.17684	2.33299,
5	7.80	10	1.99011	2.19331,
6	0.30	1	3.34111	3.23138,
6	0.45	2	3.22376	3.14164,
6	0.80	3	3.16021	3.08666,
6	1.80	4	2.93201	2.90894,
6	2.80	5	2.71902	2.73100,
6	3.80	6	2.54906	2.60510,
6	4.80	7	-999	2.59870,
6	5.80	8	2.26930	2.44538,
6	6.80	9	2.10675	2.22306,
6	7.80	10	1.97775	2.15739,
7	0.30	1	2.48738	2.34182,
7	0.45	2	2.41149	2.30957,
7	0.80	3	2.29184	2.19497,
7	1.80	4	2.10493	2.05838,
7	2.80	5	1.90630	1.89827,
7	3.80	6	1.71684	1.74078,
7	4.80	7	1.57067	1.62261,
7	5.80	8	1.38783	1.46549,
7	6.80	9	1.26306	1.36204,
7	7.80	10	1.12656	1.25589,
8	0.30	1	3.63611	3.47015,
8	0.80	2	3.32258	3.22538,
8	1.80	3	2.89675	2.88807,
8	2.80	4	2.50003	2.54646,
8	3.80	5	2.16647	2.26991,
8	4.80	6	-999	1.94030,
8	5.80	7	1.53860	1.75135,
8	7.80	8	1.11277	1.27193,
9	0.30	1	2.64678	2.56522,
9	0.45	2	2.51201	2.46954,
9	0.80	3	2.37777	2.31604,
9	1.80	4	2.09786	2.09272,
9	2.80	5	1.84498	1.87358,
9	3.80	6	1.61762	1.68425,
9	5.80	7	1.14205	1.26456,
9	7.80	8	0.82468	0.94117,
10	0.30	1	4.15769	4.07425,
10	0.80	2	3.74063	3.66860,
10	1.80	3	3.17787	3.17309,
10	2.80	4	2.63972	2.70451,
10	3.80	5	2.15830	2.32274,
10	5.80	6	1.55962	1.69767,
10	7.80	7	1.08578	1.25006,
11	0.30	1	4.70576	4.52991,
11	0.45	2	4.58412	4.35606,
11	0.80	3	4.26297	4.12025,
11	1.80	4	3.84957	3.83240,
11	2.80	5	3.38151	3.40579,
11	3.80	6	3.02387	3.17685,
11	5.80	7	2.34690	2.49886,
11	7.80	8	1.87995	2.08678,
12	0.30	1	3.80480	3.62541,
12	0.45	2	3.59962	3.48708,
12	0.80	3	3.52828	3.32105,
12	1.80	4	3.15069	3.10180,
12	2.80	5	2.87971	2.97700,
12	3.80	6	2.58227	2.69846,

```
 12     5.80   7  2.09486    2.23495,
 12     7.80   8  1.71963    1.90039};

data=uni(rawdata);

print data;
/*-----------------------------------------------------------------*
| The user must change the following information as required.      |
|                                                                  |
| scol = the column number in the DATA matrix that contains the    |
|        subject ID.                                               |
|                                                                  |
| tcol = the column number in the DATA matrix that contains the    |
|        independent variable (usaully time or dose).              |
|                                                                  |
| ocol = the column in the DATA matrix that contains the variable  |
|        that identifies which occurance for the observation.      |
|                                                                  |
| vcol = the column in the DATA matrix that contains the variable  |
|        that identifies which multivariate variabel is being      |
|        used.                                                     |
|                                                                  |
| rcol = the column in the DATA matrix that contains the           |
|        dependent variable.                                       |
|                                                                  |
|    p = the total number of fixed effects in the model.           |
|                                                                  |
|    k = the total number of random effects in the model.          |
|                                                                  |
|    r = the number of independent variables.                      |
|                                                                  |
*-----------------------------------------------------------------*/

tcol=2; scol=1; ocol=3; vcol=4; rcol=5;
p=4; k=4; r=2;

/*-----------------------------------------------------------------*
|                                                                  |
|  The user must provide initial estimates for:                    |
|                                                                  |
|  a0 = the fixed effects (dimension p x 1).                       |
|                                                                  |
|   s = the within subject error (dimension r x r).                |
|                                                                  |
|   D = the between individualt variance/covariance matrix         |
|       (dimension k x k).                                         |
|                                                                  |
*-----------------------------------------------------------------*/

a0={
 3.774067521059,
-0.118765830373,
 3.6035332163139,
-0.098022585194};


s={
0.0015215081822 0.0013622775346,
0.0013622775346  0.002020751216};

D={
 0.5005639015672 -0.004352555125  0.473558897874 -0.003949024314,
-0.004352555125 0.0009373859838 -0.004161697689 0.0008683960924,
 0.473558897874 -0.004161697689 0.4485704542735 -0.003789586783,
-0.003949024314 0.0008683960924 -0.003789586783 0.0008047989647};
```

```
/*-----------------------------------------------------------------------*
|                                                                       |
|  The user must specify the following:                                 |
|                                                                       |
|  method = maximum likelihood or restricted maximum likelihood (ML or RML).
|                                                                       |
|  approc = approximation method (LB=Lindstrom and Bates SB=Sheiner and Beale).
|                                                                       |
|  maxiter1 = the maximum number of iterations for the nonlinear loop.  |
|                                                                       |
|  maxiter2 = the maximum number of iterations for the EM algorithm loop.
|                                                                       |
|  converge = the convergence criteria for change -2lnLikelihood        |
|                                                                       |
*-----------------------------------------------------------------------*/

method='ml'; approx='LB'; maxiter1=10000; maxiter2=100; converge=0.0001;

b0=repeat(0,12,k);  /* This initializes b0 */
bup=b0;             /* This initializes the updated bi */
finish;


/*-----------------------------------------------------------------------*
|                                                                       |
|  The function "uni" converts a multivariate type of data set to a univariate
|  type of data set.                                                    |
|                                                                       |
*-----------------------------------------------------------------------*/
start uni(data);
  flag=-999;
  r=ncol(data)-3; print r;
  m=nrow(data); print m;
  c=m*r;
  oldi=0;
  datanew=repeat(0,c,5);
do i=1 to m;
 do j=1 to r;
   k=3+j;
   ii=(i-1)*r+j;
   datanew[ii,]=data[i,1:3] || j || data[i,k];
 end;
end;
do i=1 to c;
  if datanew[i,5]^=flag then do;
     if i=1 then do;
        data2=datanew[i,];
       end;
       else do;
        data2=data2//datanew[i,];
       end;
 end;
end;
return(data2);
finish;


/*   SAMPLE SIZE DETERMINATION   */

start size;
  n=nrow(data); ni=1; m=1; nij=1; no=1; qi=1; qimax=0;
  lastsid=data[1,scol]; lastoid=data[1,ocol];
  do o=2 to n; sid=data[o,scol]; oid=data[o,ocol];
    if sid=lastsid then ni=ni+1;
    else do;
      if m=1 then NNi=ni; else NNi=NNi||ni;
      if m=1 then QQi=qi; else QQi=QQi||qi;
      ni=1; m=m+1; lastsid=sid; if qi > qimax then qimax=qi;
      qi=0;
    end;
    if oid=lastoid then nij=nij+1;
```

```
    else do;
      if no=1 then NNNij=nij; else NNNij=NNNij||nij;
      nij=1; no=no+1; lastoid=oid;  qi=qi+1;
    end;
  end;
  NNi=NNi||ni; QQi=QQi||qi; NNNij=NNNij||nij;
  ij=0;  NNij=repeat(0,m,qimax);
  do i = 1 to m;  qi=QQi[i];
    do j = 1 to qi;  ij=ij+1;
      NNij[i,j]=NNNij[ij];
    end;
  end;
nc=0;
/*nc=2#m;*/
  n=n+nc;
  free NNNij;
  print, n m p k no r; print NNi; print QQi; print NNij;
finish;

/* NONLINEAR LEAST SQUARES LOOP */

start nonlin;
  m2lnLik=1000;
  m2lnLik1=0;
 do iter=1 to maxiter1 while (ABS(m2lnLik-m2lnLik1)>converge);
  m2lnLik1=m2lnLik;
  run em;
  dela=a0-a;
  a0=a;
  print iter m2lnLik method approx;
  print a0 dela;
  print s D;
  print b0;
 end;
finish;


/*   E-M ALGORITHM FOR LAIRD-WARE MODEL   */

start em; m2lnL=1000; m2lnRL=1000; change=1000;
  do subiter=1 to maxiter2 while(change>converge);
    last=0; XPX=0; XPy=0; yPy=0; logdetV=0;
    do i = 1 to m;
      run subject;
      XPX=XPX+Xi`*Wi*Xi; XPy=XPy+Xi`*Wi*yi;
      yPy=yPy+yi`*Wi*yi; logdetV=logdetV+log(det(Vi));
    end;
    invXPX=inv(XPX); a=invXPX*XPy;
    m2lnL0=m2lnL; m2lnRL0=m2lnRL;
    rss=yPy-a`*XPy; object=logdetV+rss;
    constant=n#log(2#3.14159);
    m2lnL=constant+object; m2lnRL=m2lnL+log(det(XPX));
    dm2lnL=abs(m2lnL-m2lnL0); dm2lnRL=abs(m2lnRL-m2lnRL0);
    if method='ml' then change=dm2lnL;
    if method='ml' then m2lnLik=m2lnL;
    if method='rml' then change=dm2lnRL;
    if method='rml' then m2lnLik=m2lnRL;
    last=0; T1=0; T2=0;
    print subiter method approx m2lnLik change;

    do i = 1 to m;
      run subject;
      ei=Ri*Ui`*Wi*(yi-Xi*a); bi=D*Zi`*Wi*(yi-Xi*a); bup[i,]=bi`;
      if method='rml' then Wi=Wi-Wi*Xi*invXPX*Xi`*Wi;
      T2=T2+D-D*Zi`*Wi*Zi*D+bi*bi`;
      T1star=Ri-Ri*Ui`*Wi*Ui*Ri+ei*ei`;
      do j = 1 to qi;
        firstj=(j-1)#r+1;  lastj=r#j;
        T1=T1+t1star[firstj:lastj,firstj:lastj];
      end;
    end;
```

```
    D=T2/m; S=T1/no;
* S[1,2]=0;
* S[2,1]=0;
  b0=bup;
  end;
finish;


/*   MODEL SPECIFICATION FOR SUBJECT i   */
start subject;
  ni=NNi[i]; qi=QQi[i];
  do j = 1 to qi;
    nij=NNij[i,j]; first=last+1; last=last+nij;
    if approx='LB' then bi0=b0[i,]`;
    if approx='SB' then bi0=repeat(0,k,1);
    Ai=I(p); Bi=I(k);
    ci0=Ai*a0+Bi*bi0;
    c1=ci0[1]; c2=ci0[2]; c3=ci0[3]; c4=ci0[4];

    run model;
    if j = 1 then do;
      yi=yij;  Xi=Xij;  Zi=Zij; Ui=Uij;
    end;
    else do;
      yi=yi//yij;  Xi=Xi//Xij;  Zi=Zi//Zij;
      Ui=block(Ui,Uij);
    end;
  end;
  Ri=I(qi)@S; Vi=Ui*Ri*Ui`+Zi*D*Zi`; Wi=inv(Vi);
finish;

/*   MODEL SPECIFICATION FOR SUBJECT i ON OCCASION j */

/*-------------------------------------------------------------------------------*
|                                                                               |
|  The user must customize this part of the program to fit their specific nonlinear |
|  functions.                                                                   |
|                                                                               |
*-------------------------------------------------------------------------------*/

  start model;
    yyij=data[first:last,rcol];
    fij=yyij;                        /* intialize fij */
    Xij=repeat(0,nij,p);
    Uij=repeat(0,nij,r);
    do l = 1 to nij;  o = first+l-1;
      var=data[o,vcol];
      if var=1 then fij[l,]=c1*exp(c2*data[o,tcol]);
      if var=2 then fij[l,]=c3*exp(c4*data[o,tcol]);
      if var=1 then Xij[l,]=exp(c2*data[o,tcol])||c1*data[o,tcol]*exp(c2*data[o,tcol])||{0 0};
      if var=2 then Xij[l,]={0 0}||exp(c4*data[o,tcol])||c3*data[o,tcol]*exp(c4*data[o,tcol]);
      U1=data[o,vcol]//{1,2}; U2=design(U1); Uij[l,]=U2[1,];
    end;
    Zij=Xij;
    yij=yyij-fij+Xij*a0+Zij*bi0;
 finish;

/*   FISHERS INFORMATION MATRIX ON VARIANCE PARAMETERS  */
start Fisher;
  sumkint=(k#(k+1)/2);  sumrint=(r#(r+1)/2);
  q=sumkint+sumrint; THETA=repeat(0,q,1);
  do j = 1 to q;
    if j <= sumkint then do;
      rr=k; jj=j; run unvech; t=tt; u=uu; THETA[j]=D[t,u];
    end;
    else do;
       rr=r; jj=j-sumkint; run unvech; t=tt; u=uu;
       THETA[j]=S[t,u];
    end;
  end;
  F=repeat(0,q,q);last=0;
```

```
  do i = 1 to m;
    run subject;
    do j = 1 to q;
      if j <= sumkint then do;
        rr=k; jj=j; run unvech; t=tt; u=uu;
        Dtu=repeat(0,k,k);  Dtu[t,u]=1;  Dtu[u,t]=1;
        Vij=Zi*Dtu*Zi`;
      end;
      else do;
        rr=r; jj=j-sumkint; run unvech; t=tt; u=uu;
        Dtu=repeat(0,r,r);  Dtu[t,u]=1;  Dtu[u,t]=1;
        Vij=Ui*(I(qi)@Dtu)*Ui`;
      end;
      do jp = 1 to q;
        if jp <= sumkint then do;
          rr=k; jj=jp; run unvech; tp=tt; up=uu;
          Dtu=repeat(0,k,k);  Dtu[tp,up]=1;  Dtu[up,tp]=1;
          Vijp=Zi*Dtu*Zi`;
        end;
        else do;
          rr=r; jj=jp-sumkint; run unvech;  tp=tt; up=uu;
          Dtu=repeat(0,r,r);  Dtu[tp,up]=1;  Dtu[up,tp]=1;
          Vijp=Ui*(I(qi)@Dtu)*Ui`;
        end;
        F[j,jp]=F[j,jp]+trace(Wi*Vij*Wi*Vijp)/2;
      end;
    end;
  end;
  vTHETA=inv(F); print, F; print, THETA vTHETA;
finish;

/*   SUBSCRIPTS FOR rr x rr MATRIX FROM VECH SUBSCRIPT  */
start unvech;
  jjj=0;
  do ttt = 1 to rr;
    do uuu = ttt to rr;
      jjj=jjj+1;
      if jjj=jj then do;
        tt=ttt; uu=uuu;
      end;
    end;
  end;
finish;

start parmcorr;

rho=theta[4]/(sqrt(theta[1])*sqrt(theta[10]));

drhot1=-theta[4]/(2*sqrt(theta[1]**3)*sqrt(theta[10]));

drhot4=1/(sqrt(theta[1])*sqrt(theta[10]));

drhot10=-theta[4]/(2*sqrt(theta[1])*sqrt(theta[10]**3));

drhodt= drhot1||{0 0}||drhot4||{0 0 0 0 0}||drhot10||{0 0 0};

Vrho=drhodt*vtheta*drhodt`;

SErho=sqrt(Vrho);

print rho Vrho SErho;

finish;
run init; run size; run nonlin; run fisher; run parmcorr;
```

APPENDIX B

SAS/IML Output from Implementation of MNLMEM on the Doubly Labeled Water Data Set

```
        SUBITER METHOD APPROX         M2LNLIK          CHANGE
              1 ml     LB     -625.2695978058 1625.2695978058


        SUBITER METHOD APPROX         M2LNLIK          CHANGE
              2 ml     LB     -625.2696011315 3.3257138057E-6


           ITER         M2LNLIK METHOD APPROX
              2 -625.2696011315 ml     LB


             A0            DELA
3.7740674495849 -2.858395964E-8
-0.118765801476 1.1759749605E-8
 3.603533767924 -2.171445592E-7
-0.098022710907 3.8649749928E-9


             S                                    D
0.0015215111039 0.0013622765505 0.5005639256633 -0.004352578385  0.473558071447 -0.003948873815
0.0013622765505  0.002020757796 -0.004352578385 0.0009373842937 -0.004161733707 0.0008683992088
                                 0.473558071447 -0.004161733707 0.4485687619091 -0.003789455142
                                -0.003948873815 0.0008683992088 -0.003789455142 0.0008048061691


             B0
-0.266382253316 0.0309580593609 -0.226134244508 0.0280330611766
1.1232677056215 0.0108513894844 1.0540430778556 0.0104345856029
 -0.46761894643 0.0031042939952 -0.426795566351 0.0024395748218
0.3294988141049 -0.026092596711 0.2825006905455 -0.023439212358
0.2878382244479 0.0263376461123 0.2593326270154 0.0247293651261
-0.428102816537  0.050819757926 -0.384938895537 0.0465317101601
-1.243907039587 0.0162931262984  -1.21098465773 0.0156711591258
0.0270432885607 -0.034271586716 0.0243405862641  -0.03170161512
-1.050269776513  -0.02752976375 -0.985471790299 -0.025855083381
0.5856345677921  -0.06076834238 0.5850842543517 -0.056897170009
1.0363183783957 -0.004322158476 0.9753086370103 -0.003716347299
0.0666798534613  0.014620174856 0.0537152813834 0.0137699721544

475                              run fisher;
```

F

19053953.806116 425859616.13476 -40171045.15763 -461216078.5001 2514149298.7284 -448876590.2029
: -5444645565.826 21172942.676956 486144170.35955 2947735441.0672 4971559.4917801 -9983003.047661
: 5035142.3461646

425859616.13476 16062681527.238 -896862554.7598 -17347157036.86 133719016358.89 -16919344597.79
: -289036974574.3 472198899.55471 18272292933.824 156189572852.45 184139130.0443 -370138335.3104
: 187238048.18543

-40171045.15763 -896862554.7598 84741036.698842 971563309.98067 -5289057051.768 945871108.62591
: 11456594038.52 -44690411.68767 -1024654141.43 -6204019827.414 -10470578.27742 21034036.92645
: -10613066.73416

-461216078.5001 -17347157036.86 971563309.98067 18761752647.714 -144259289993.6 18274893068.951
: 312116808764.15 -511655481.4446 -19765135478.63 -168822512660 -198873272.4628 400041316.19596
: -202537586.3622

2514149298.7284 133719016358.89 -5289057051.768 -144259289993.6 1801676542907.4 -140646310203.4
: -3887624190760 2781669528.3777 151732586428.69 2097161516442.4 3171287199.264 -6435316189.919
: 3283021171.587

-448876590.2029 -16919344597.79 945871108.62591 18274893068.951 -140646310203.4 17835344046.567
: 304076818845.69 -498282931.9899 -19264223310.3 -164352489204.7 -193475352.2735 388908054.78238
: -196721540.9979

-5444645565.826 -289036974574.3 11456594038.52 312116808764.15 -3887624190760 304076818845.69
: 8396215542413.4 -6026727233.563 -328357500276.5 -4533383489905 -6836234513.399 13888890257.754
: -7094258568.692

21172942.676956 472198899.55471 -44690411.68767 -511655481.4446 2781669528.3777 -498282931.9899
: -6026727233.563 23582378.650758 539919372.97908 3264359765.614 5513040.9501338 -11079684.1959
: 5592612.2352753

486144170.35955 18272292933.824 -1024654141.43 -19765135478.63 151732586428.69 -19264223310.3
: -328357500276.5 539919372.97908 20838086989.487 177645700774.32 208958586.4523 -420331236.29
: 212797632.50339

2947735441.0672 156189572852.45 -6204019827.414 -168822512660 2097161516442.4 -164352489204.7
: -4533383489905 3264359765.614 177645700774.32 2449926785540 3684183419.2407 -7493860134.605
: 3832490717.0588

4971559.4917801 184139130.0443 -10470578.27742 -198873272.4628 3171287199.264 -193475352.2735
: -6836234513.399 5513040.9501338 208958586.4523 3684183419.2407 110283558.81409 -153799177.8674
: 54836246.327173

-9983003.047661 -370138335.3104 21034036.92645 400041316.19596 -6435316189.919 388908054.78238
: 13888890257.754 -11079684.1959 -420331236.29 -7493860134.605 -153799177.8674 279758784.32373
: -120829835.6828

5035142.3461646 187238048.18543 -10613066.73416 -202537586.3622 3283021171.587 -196721540.9979
: -7094258568.692 5592612.2352753 212797632.50339 3832490717.0588 54836246.327173 -120829835.6828
: 66891173.581819

THETA

0.5005639256633
-0.004352578385
0.473558071447
-0.003948873815
0.0009373842937
-0.004161733707
0.0008683992088
0.4485687619091
-0.003789455142
0.0008048061691
0.0015215111039
0.0013622765505
0.002020757796

```
          VTHETA
  0.0418764612216 -0.000367689814 0.0396129618901 -0.000333521267 3.2285130827E-6 -0.000347815469
: 2.9284909031E-6 0.0374718088492 -0.000315493798 2.6563496966E-6 -2.737123081E-8 -2.443281883E-8
: -2.175022377E-8

  -0.000367689814 0.0000410623089 -0.000349490977 0.0000379906357 -6.927349111E-7  0.000038857527
: -6.348945433E-7 -0.000332185259 0.0000359505162 -5.818219793E-7 1.9112882244E-9 1.7169473206E-9
: 1.5411583331E-9

  0.0396129618901 -0.000349490978 0.0375089513659 -0.000318468333 3.0834255558E-6 -0.000330926511
: 2.8096724763E-6 0.0355166586478  -0.00030155038 2.5601696158E-6 -2.376434425E-8 -2.798458396E-8
: -3.102951731E-8

  -0.000333521267 0.0000379906357 -0.000318468333 0.0000352777936 -6.414288272E-7 0.0000359632906
: -5.890004582E-7 -0.000304068249 0.0000333946554 -5.407770563E-7 1.6967343358E-9 1.9624500445E-9
: 2.1558225568E-9

  3.2285130654E-6 -6.927349107E-7 3.0834255394E-6 -6.414288269E-7 1.4864121172E-7 -6.616047188E-7
: 1.3763236095E-7 2.9448584514E-6  -6.12604209E-7 1.2743885917E-7 -2.73268526E-10 -2.47672873E-10
: -2.24792126E-10

  -0.000347815469  0.000038857527 -0.000330926511 0.0000359632906 -6.616047192E-7 0.0000368412313
: -6.091040836E-7 -0.000314853073 0.0000340970625 -5.607511094E-7  1.630878679E-9 1.8852064023E-9
: 2.0752015801E-9

  2.9284908872E-6  -6.34894543E-7 2.8096724612E-6 -5.890004578E-7 1.3763236095E-7 -6.091040833E-7
: 1.2768082439E-7 2.6956179185E-6 -5.650695032E-7 1.1844842834E-7 -2.35601089E-10 -2.68710405E-10
: -2.93813726E-10

  0.0374718088491  -0.00033218526 0.0355166586478 -0.000304068249 2.9448584669E-6 -0.000314853074
: 2.6956179328E-6 0.0336635226762 -0.000288203171 2.4674806511E-6 -2.056338646E-8  -3.01784813E-8
: -4.407776515E-8

  -0.000315493798 0.0000359505162  -0.00030155038 0.0000333946554 -6.126042093E-7 0.0000340970625
: -5.650695035E-7 -0.000288203171   0.00003167267 -5.211927331E-7 1.4448685267E-9 2.0504433556E-9
: 2.9095258218E-9

  2.6563496818E-6  -5.81821979E-7 2.5601696018E-6  -5.40777056E-7 1.2743885917E-7 -5.607511091E-7
: 1.1844842834E-7 2.4674806379E-6 -5.211927328E-7 1.1009218227E-7 -2.03109641E-10 -2.79464307E-10
: -3.85887016E-10

  -2.737123077E-8 1.9112882245E-9 -2.376434421E-8 1.6967343359E-9 -2.73268527E-10 1.6308786791E-9
: -2.35601089E-10 -2.056338643E-8 1.4448685268E-9 -2.03109641E-10 6.0263739478E-8  5.363432531E-8
: 4.7538350567E-8

  -2.443281879E-8 1.7169473208E-9 -2.798458392E-8 1.9624500447E-9 -2.47672873E-10 1.8852064025E-9
: -2.68710405E-10 -3.017848126E-8 2.0504433558E-9 -2.79464307E-10  5.363432531E-8 6.4057431628E-8
: 7.1491086419E-8

  -2.175022373E-8 1.5411583335E-9 -3.102951726E-8 2.1558225572E-9 -2.24792127E-10 2.0752015804E-9
: -2.93813727E-10  -4.40777651E-8 2.9095258222E-9 -3.85887016E-10 4.7538350567E-8 7.1491086419E-8
: 1.0655848237E-7


475                                          run parmcorr;

           RHO            VRHO           SERHO
-0.196742340643 0.0778412021067 0.2790003621982
```

APPENDIX C

MATLAB Program Implementing MNLMEM on the Glucose Tolerance Test Data Set

```
% /*-----------------------------------------------------------*
% |      MULTIVARIATE NONLINEAR MIXED EFFECTS MODEL           |
% |       Jim Rutledge's version of 8 August 1993             |
% |                                                           |
% |   This algorithm is based on the maximum likelihood and   |
% |   restricted maximum likelihood algorithms proposed by    |
% |   Hocking in his book The Analysis of Linear Models (ch 8).|
% |                                                           |
% *-----------------------------------------------------------*/

% NOTE:  This version of the program uses convergence of -2 log
%        likelihood for the inside Hocking loop.  The convergence
%        of the non-linear parameters is used in the outer Newton-
%        Raphson loop.

% DATA:  Ron Gotlin's data from a glucose tolerance test.
%        This program models phosphate and insulin levels for
%        the control group.  The time lag between max insulin
%        and min phosphate is estimated.

% DATE:  This program was revised on 20 November 1994.

% The user must supply Ao (an initial guess) for the fixed effects.
% The individual least squares solution may provide a good guess.
% The user must specify an initial guess for the SI (the variance
% components).

format long
clear
tic
% /**** INITIAL VALUE ****/
Ao =[
   1.300050028277
   0.661658368697
   61.517432378772
  -0.501793228314
   0.010423565490
   68.662393142131
   265.152281605558]

SI=[
   2.86769926486938
   1.81292281093675
   1.28880294909515
   1.02797490997373
   0.17566081602352
  -0.35613748135715
  -1.45150651269620
  -2.24218716017727
   0.30043737616578
   0.15568697682150
   0.09836749207982
   0.09199411143703
   0.12034755996616
   0.15012514728637
   0.00927951563554
   0.02270942515335
  -0.06427436936667]

n=208;
p=max(size(Ao));
k=max(size(SI));
```

```
Aoold=zeros(size(Ao));
Aold=zeros(size(Ao));
RO=zeros(k,1);
W=eye(k);
m2lnL=1000;
m2lnRL=1000;
diffout=1000;
diffin=1000;
converge=0.00001;
method='ml ';
ocount=0;
incount=0;

%************************************************************************
% The user must customize the model subroutine.  The model used is:
%
%                                   Y=X*ALPHA + e
%
% Where the V(e)=SI[1]*V1+ ... +SI[k]*Vk.
%
% The user must specify the n x n variance matrices V1,V2,...,Vk, the
% n x p design matrix X, and the n x 1 data matrix Y.
%************************************************************************


%**** MODEL SPECIFICATION ****

DATA=[
 1 0 .001  75  24 4.3
 1 0  30 142  97  3.3
 1 0  60 132 100  3.0
 1 0  90 100  87  2.6
 1 0 120  84  71  2.2
 1 0 180  80  29  2.5
 1 0 240  76  20  3.4
 1 0 300  72  20  4.4
 2 0 .001  69  30 3.7
 2 0  30 139  55  2.6
 2 0  60 142  90  2.6
 2 0  90 131  70  1.9
 2 0 120 111  42  2.9
 2 0 180  96  36  3.2
 2 0 240  80  20  3.1
 2 0 300  73  28  3.9
 3 0 .001  79  10 4.0
 3 0  30 160  60  4.1
 3 0  60 150  72  3.1
 3 0  90 122  54  2.3
 3 0 120 102  36  2.9
 3 0 180  84  30  3.1
 3 0 240  70  22  3.9
 3 0 300  76  25  4.0
 4 0 .001  64  11 3.6
 4 0  30 104  45  3.0
 4 0  60  79  28  2.2
 4 0  90 121  50  2.8
 4 0 120 100  32  2.9
 4 0 180  74  11  3.9
 4 0 240  70   8  3.8
 4 0 300  71  15  4.0
 5 0 .001  90  20 4.1
 5 0  30 157  61  3.8
 5 0  60 117  95  2.1
 5 0  90 118  70  3.0
 5 0 120  89  50  3.6
 5 0 180  71  32  3.4
 5 0 240  88  25  3.6
 5 0 300  82  10  3.7
 6 0 .001  55  13 3.8
 6 0  30 102  72  2.2
 6 0  60  78  68  2.0
```

```
 6 0  90 111  48  2.6
 6 0 120  67  40  3.8
 6 0 180  70  28  3.6
 6 0 240  74  10  3.0
 6 0 300  68  15  3.5
 7 0 .001  77  15 3.8
 7 0  30 125  80  3.0
 7 0  60 104  84  2.4
 7 0  90 101  64  2.5
 7 0 120 104  38  3.1
 7 0 180  69  22  3.4
 7 0 240  89  16  3.5
 7 0 300  75  16  3.7
 8 0 .001  88  10 4.4
 8 0  30 188  67  3.9
 8 0  60 170  75  2.8
 8 0  90 104  60  2.1
 8 0 120 102  44  3.6
 8 0 180  76  18  3.8
 8 0 240  80  12  4.0
 8 0 300  62  14  3.9
 9 0 .001  84  35 5.0
 9 0  30 144  84  4.0
 9 0  60 129  90  3.4
 9 0  90 126 102  3.4
 9 0 120 112  87  3.3
 9 0 180  94  47  3.6
 9 0 240  69  32  4.0
 9 0 300  78  15  4.3
10 0 .001  86  11 3.7
10 0  30 157  45  3.1
10 0  60 117  28  2.9
10 0  90 116  50  2.2
10 0 120  89  32  1.5
10 0 180  71  11  2.3
10 0 240  88  25  2.7
10 0 300  82  16  2.8
11 0 .001  69  24 3.7
11 0  30 120  97  2.6
11 0  60 119 100  2.6
11 0  90  99  87  2.3
11 0 120 100  71  2.9
11 0 180  57  29  2.2
11 0 240  75  20  3.1
11 0 300  79  20  3.9
12 0 .001  76  10 4.4
12 0  30 128  33  3.7
12 0  60  90  39  3.1
12 0  90  50  25  3.2
12 0 120  68  18  3.7
12 0 180  50  10  4.3
12 0 240  64  10  3.9
12 0 300  72  10  4.8
13 0 .001  80   4 4.7
13 0  30 138  91  3.1
13 0  60 127  76  3.2
13 0  90 100  31  3.3
13 0 120  96  30  3.2
13 0 180  62  10  4.2
13 0 240  76  10  3.7
13 0 300  80  10  4.3
];
subV1=[
1 0 0 0 0 0 0 0
0 1 0 0 0 0 0 0
0 0 1 0 0 0 0 0
0 0 0 1 0 0 0 0
0 0 0 0 1 0 0 0
0 0 0 0 0 1 0 0
0 0 0 0 0 0 1 0
0 0 0 0 0 0 0 1];
```

```
subV2=[
0 1 0 0 0 0 0 0
1 0 1 0 0 0 0 0
0 1 0 1 0 0 0 0
0 0 1 0 1 0 0 0
0 0 0 1 0 1 0 0
0 0 0 0 1 0 1 0
0 0 0 0 0 1 0 1
0 0 0 0 0 0 1 0];

subV3=[
0 0 1 0 0 0 0 0
0 0 0 1 0 0 0 0
1 0 0 0 1 0 0 0
0 1 0 0 0 1 0 0
0 0 1 0 0 0 1 0
0 0 0 1 0 0 0 1
0 0 0 0 1 0 0 0
0 0 0 0 0 1 0 0];

subV4=[
0 0 0 1 0 0 0 0
0 0 0 0 1 0 0 0
0 0 0 0 0 1 0 0
1 0 0 0 0 0 1 0
0 1 0 0 0 0 0 1
0 0 1 0 0 0 0 0
0 0 0 1 0 0 0 0
0 0 0 0 1 0 0 0];

subV5=[
0 0 0 0 1 0 0 0
0 0 0 0 0 1 0 0
0 0 0 0 0 0 1 0
0 0 0 0 0 0 0 1
1 0 0 0 0 0 0 0
0 1 0 0 0 0 0 0
0 0 1 0 0 0 0 0
0 0 0 1 0 0 0 0];

subV6=[
0 0 0 0 0 1 0 0
0 0 0 0 0 0 1 0
0 0 0 0 0 0 0 1
0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0
1 0 0 0 0 0 0 0
0 1 0 0 0 0 0 0
0 0 1 0 0 0 0 0];

subV7=[
0 0 0 0 0 0 1 0
0 0 0 0 0 0 0 1
0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0
1 0 0 0 0 0 0 0
0 1 0 0 0 0 0 0];

subV8=[
0 0 0 0 0 0 0 1
0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0
1 0 0 0 0 0 0 0];
```

```
V1=sparse([  kron(eye(13),subV1) zeros(104)
             zeros(104)          zeros(104)]);

V2=sparse([  kron(eye(13),subV2) zeros(104)
             zeros(104)          zeros(104)]);

V3=sparse([  kron(eye(13),subV3) zeros(104)
             zeros(104)          zeros(104)]);

V4=sparse([  kron(eye(13),subV4) zeros(104)
             zeros(104)          zeros(104)]);

V5=sparse([  kron(eye(13),subV5) zeros(104)
             zeros(104)          zeros(104)]);

V6=sparse([  kron(eye(13),subV6) zeros(104)
             zeros(104)          zeros(104)]);

V7=sparse([  kron(eye(13),subV7) zeros(104)
             zeros(104)          zeros(104)]);

V8=sparse([  kron(eye(13),subV8) zeros(104)
             zeros(104)          zeros(104)]);

V9=sparse([  zeros(104)          zeros(104)
             zeros(104)          kron(eye(13),subV1)]);

V10=sparse([  zeros(104)          zeros(104)
              zeros(104)          kron(eye(13),subV2)]);

V11=sparse([  zeros(104)          zeros(104)
              zeros(104)          kron(eye(13),subV3)]);

V12=sparse([  zeros(104)          zeros(104)
              zeros(104)          kron(eye(13),subV4)]);

V13=sparse([  zeros(104)          zeros(104)
              zeros(104)          kron(eye(13),subV5)]);

V14=sparse([  zeros(104)          zeros(104)
              zeros(104)          kron(eye(13),subV6)]);

V15=sparse([  zeros(104)          zeros(104)
              zeros(104)          kron(eye(13),subV7)]);

V16=sparse([  zeros(104)          zeros(104)
              zeros(104)          kron(eye(13),subV8)]);

V17=sparse([  zeros(104)          kron(eye(13),subV1)
              kron(eye(13),subV1) zeros(104)]);


%/**** NON-LINEAR ***********/

while diffout > converge;
  a11=Ao(1,1);
  a12=Ao(2,1);
  a13=Ao(3,1);
  a21=Ao(4,1);
  a22=Ao(5,1);
  a23=Ao(6,1);
  a24=Ao(7,1);

  x1=DATA(1:104,3);
  x2=x1;

  yy1=DATA(1:104,5)./10;
  yy2=DATA(1:104,6)./1;
```

```
% The user must specify the form of the nonlinear functions
% to be fit to the data.  The derivatives also need to be
% specified.


  f1=a11+x1.^a12.*exp((-1/a13).*x1);
  f2=a21+a22.*(x2+a23)+a24./(x2+a23);

  da11=ones(size(x1));
  da12=x1.^a12.*log(x1).*exp((-1/a13).*x1);
  da13=x1.^a12.*(x1./(a13^2)).*exp((-1/a13).*x1);
  da21=ones(size(x2));
  da22=x2+a23;
  da23=a22-a24./((x2+a23).^2);
  da24=1./(x2+a23);

  y1=yy1-f1;
  y2=yy2-f2;

  X=sparse([da11 da12 da13 zeros(104,4);
            zeros(104,3) da21 da22 da23 da24]);

  Y=[y1;y2];


%/**** HOCKING ALGORITHM ****/

initer=1;
maxiter=10;
while diffin > converge;
incount=incount+1

% The user must specify the form for V.

V=sparse(SI(1,1)*V1+SI(2,1)*V2+SI(3,1)*V3+SI(4,1)*V4+SI(5,1)*V5+SI(6,1)*V6+...
         SI(7,1)*V7+SI(8,1)*V8+SI(9,1)*V9+SI(10,1)*V10+SI(11,1)*V11+SI(12,1)*V12+...
         SI(13,1)*V13+SI(14,1)*V14+SI(15,1)*V15+SI(16,1)*V16+SI(17,1)*V17);
IV=inv(V);
if method=='rml';
  IV=inv(V)-(inv(V)*X*inv(X'*inv(V)*X)*X'*inv(V));
end;
XPX=X'*inv(V)*X;
XPY=X'*inv(V)*Y;
YPY=Y'*inv(V)*Y;
A=inv(XPX)*XPY;

for i=1:k;
  for j=1:k;
   eval(['W(i,j)=trace(IV*V' int2str(i) '*IV*V' int2str(j) ');']);
  end;
end;

OMEGA=W;

YmXAT=(Y-X*A)';
for i=1:k;
  eval(['RO(i,1)=YmXAT*IV*V' int2str(i) '*IV*(Y-X*A);']);
end;

RHO=RO;
SI=inv(OMEGA)*RHO;

m2lnLold=m2lnL;
m2lnRLol=m2lnRL;
rss=(Y-X*A)'*inv(V)*(Y-X*A);
detV=det(V)
object=log(detV)+rss;
 n=max(size(Y));
constant=n*log(2*3.14159265);
```

```
m2lnL=constant+object;
m2lnRL=m2lnL+log(det(X'*inv(V)*X));
dm2lnL=abs(m2lnL-m2lnLold);
dm2lnRL=abs(m2lnRL-m2lnRLol);

if method=='rml';
  diffin=dm2lnRL;
end;
if method=='ml ';
  diffin=dm2lnL;
end;
incount, m2lnL, m2lnRL
initer=initer+1;
end;
m2lnL=1000;
diffin=1000;
incount=0
diffout=norm(Ao-Aoold);
Aoold=Ao;
Ao=Ao+A;

ocount=ocount+1
ocount, m2lnL, m2lnRL, diffin, diffout, A, Ao, SI
end
Valpha=inv(X'*inv(V)*X)

% This part of the program estimates a nonlinear funtion of
% the parameters.  The user can change this if desired.

maxIn=Ao(2,1)*Ao(3,1)
minPhos=-(sqrt(Ao(5,1))*Ao(6,1)-sqrt(Ao(7,1)))/sqrt(Ao(5,1))
g=minPhos-maxIn
dgda1=0;
dgda2=Ao(3,1);
dgda3=Ao(2,1);
dgda4=0;
dgda5=-sqrt(Ao(7,1))/2*Ao(5,1)^1.5;
dgda6=-1;
dgda7=1/(2*sqrt(Ao(5,1)*Ao(7,1)));
dgda=[dgda1 dgda2 dgda3 dgda4 dgda5 dgda6 dgda7]
Vg=dgda*Valpha*dgda'
seVg=sqrt(Vg)

time=toc
```

APPENDIX D

MATLAB Output from Implementation of MNLMEM on the Glucose Tolerance Test Data Set

```
m2lnL =

    4.408879122907945e+002

m2lnRL =

    4.577920807688517e+002

incount =

     2

detV =

    1.319844279947076e-065

incount =

     2

m2lnL =

    4.408879122907933e+002

m2lnRL =

    4.577920810058294e+002

incount =

     0

ocount =

    21

ocount =

    21

m2lnL =

        1000

m2lnRL =

    4.577920810058294e+002

diffin =
```

```
        1000

diffout =

    9.988600625851892e-006


A =

  1.0e-005 *

  -0.00015770049412
  -0.00000268487961
   0.00351017690059
   0.00447126327161
  -0.00000587159994
  -0.09003748241892
  -0.56728130089824


Ao =

  1.0e+002 *

   0.01382819282386
   0.00649626764841
   0.63409689486964
  -0.00504904479497
   0.00010352797906
   0.71657198839682
   2.67872230307720


SI =

   2.86769905524166
   1.81292151927484
   1.28880009445575
   1.02797037782445
   0.17565478144361
  -0.35614394284830
  -1.45151164988108
  -2.24219322026139
   0.30043743696911
   0.15568701638798
   0.09836757091213
   0.09199436007768
   0.12034788821133
   0.15012513892531
   0.00927940250155
   0.02270931858372
  -0.06427482425216

Valpha =

  1.0e+004 *

   (1,1)       0.00000137321900
   (2,1)       0.00000000978822
   (3,1)       0.00000957329433
   (4,1)       0.00000206142877
   (5,1)      -0.00000000407615
   (6,1)      -0.00002480274494
   (7,1)      -0.00021688730795
   (1,2)       0.00000000978822
   (2,2)       0.00000005377500
   (3,2)      -0.00001107432012
   (4,2)       0.00000078294124
   (5,2)      -0.00000000108237
   (6,2)      -0.00001666455940
```

```
    (7,2)      -0.00009797207142
    (1,3)       0.00000957329433
    (2,3)      -0.00001107432012
    (3,3)       0.00353370904076
    (4,3)      -0.00017667088905
    (5,3)       0.00000026139787
    (6,3)       0.00344742212382
    (7,3)       0.02136120121562
    (1,4)       0.00000206142877
    (2,4)       0.00000078294124
    (3,4)      -0.00017667088905
    (4,4)       0.00014508210056
    (5,4)      -0.00000020832449
    (6,4)      -0.00269151347156
    (7,4)      -0.01728645642273
    (1,5)      -0.00000000407615
    (2,5)      -0.00000000108237
    (3,5)       0.00000026139787
    (4,5)      -0.00000020832449
    (5,5)       0.00000000032157
    (6,5)       0.00000367160241
    (7,5)       0.00002411987429
    (1,6)      -0.00002480274494
    (2,6)      -0.00001666455940
    (3,6)       0.00344742212382
    (4,6)      -0.00269151347156
    (5,6)       0.00000367160241
    (6,6)       0.05394597315306
    (7,6)       0.33394390654825
    (1,7)      -0.00021688730795
    (2,7)      -0.00009797207142
    (3,7)       0.02136120121562
    (4,7)      -0.01728645642274
    (5,7)       0.00002411987429
    (6,7)       0.33394390654824
    (7,7)       2.11299726363316

maxIn =

  41.19263144096442

minPhos =

  89.19795435927152

g =

  48.00532291830710

dgda =

  Columns 1 through 4

                  0  63.40968948696425   0.64962676484061                  0

  Columns 5 through 7

  -0.00862026180057  -1.00000000000000   0.30024604083479

Vg =

    4.692892806029775e+002

seVg =

  21.66308566670449

time =

    1.969790000000000e+003
```